

Ei evaluering av talekvaliteten til den syntetiske stemmen Brage

*Med vekt på velegnethet i forhold til
formidling av studielitteratur*

Marius Olaussen



Masteroppgave i nordisk, særlig norsk, språkvitenskap

Institutt for lingvistiske og nordiske studier
Det humanistiske fakultet

UNIVERSITETET I OSLO

Høsten 2010

Ei evaluering av talekvaliteten til den syntetiske stemmen Brage

*Med vekt på velegnethet i forhold til
formidling av studielitteratur*

Marius Olaussen



Masteroppgave i nordisk, særlig norsk, språkvitenskap

Institutt for lingvistiske og nordiske studier
Det humanistiske fakultet

UNIVERSITETET I OSLO

Høsten 2010

© Marius Olaussen

2010

Ei evaluering av talekvaliteten til den syntetiske stemmen Brage. Med vekt på velegnethet i forhold til formidling av studielitteratur.

Marius Olaussen

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

Sammendrag

Syns- og lesehemmete studenter i høyere utdanning har behov for tilpasset studielitteratur. Norsk lyd- og blindeskriftbibliotek har ansvaret for å tilrettelegge og produsere denne studielitteraturen. Over halvparten av studiebøkene NLB tilrettelegger, produseres med talesyntese. For å styrke tilbud til studentene, ble NLB i 2008 bevilget én million kroner av Kunnskapsdepartementet. NLB inngikk et samarbeid med det svenske Talboks- og punktskriftsbiblioteket om å tilpasse deres ”tekst-til-tale”-system Filibuster fra svensk til norsk bokmål. Filibuster ble særlig utviklet for å tolke og behandle utfordringer som særpreger studielitteratur. Høsten 2009 ble den norske stemmen Brage lansert.

Hensikten med denne studien er å evaluere talekvaliteten til Brage med tanke på hvor velegnet stemmen var til å formidle studielitteratur. Fire forskningsspørsmål ble formulert:

- Hvordan opplever studenter med et lesehandikap talekvaliteten til Filibuster med Brage, vurdert etter sentrale kriterier innenfor evalueringsforskninga?
- Hvor tydelig er Brage sammenliknet med andre norske talesynteser?
- Hvor velegnet synes Brage å være til å formidle studielitteratur?
- Hvordan bør Filibuster med Brage videreutvikles?

Det ble gjennomført en brukerundersøkelse og en tydelighetstest (SUS-test). Utvalget til brukerundersøkelsen besto av 79 studenter. Data ble samlet inn gjennom et spørreskjema med spørsmål og svaralternativ, samt noen åpne spørsmål. Følgende evalueringskriterier ble lagt til grunn for spørreskjemaet: velegnethet, tydelighet, naturlighet, aksept og behagelighet. En SUS-test med 18 informanter ble i tillegg brukt for å måle tydelighet. Det empiriske materialet viste at studentene i hovedsak opplevde talekvaliteten til Brage som god. Brage synes å være velegnet til å produsere studielitteratur generelt sett, men må trenes ytterligere innenfor visse studieretninger. SUS-testen viste at Brage sammenliknet med to andre syntetiske stemmene, skåret jevnt over høyere enn begge disse.

Studien har gitt viktig informasjon for det videre arbeidet. Det er foreslått enkelte verktøy og tiltak for å forbedre talekvaliteten til Brage.

Forord

Jeg vil først få lov til å takke veilederen min, Janne Bondi Johannessen, for verdifulle innspill og god veiledning underveis. Takk for at jeg også fikk mulighet til å knytte dette arbeidet oppimot mastergraden min.

En stor takk til alle informanter som deltok i denne studien og som ville dele av sine erfaringer og kunnskaper. Deres bidrag er helt avgjørende for det videre arbeidet med Brage.

Hjertelig takk skal Gunhild Hovden Kvangarsnes også ha for all hjelp i forbindelse med brukerundersøkelsen og innsamling av data.

Videre vil jeg takke min overordnede, Arne Kyrkjebø, som ga meg tid til å fullføre dette arbeidet. En takk til alle mine kolleger ved Norsk lyd- og blindeskriftbibliotek, som har vært behjelpelige i løpet av denne prosessen.

Jeg vil også takke Christina Tånnander og Kåre Sjølander for at jeg har fått lov til å samarbeide med dere så lenge, og for alt jeg har lært av dere i løpet av denne tida.

Takk til Jan Erik Kaplon for din innsats med Brage.

Til slutt vil jeg takke familie og nærmeste for viktig støtte.

Marius Olaussen

Oslo, november 2010

Innholdsfortegnelse

1	Innledning	1
1.1	Tilrettelagt litteratur og talesyntese	1
1.2	Denne studiens formål og forskningsspørsmål.....	2
1.3	Oppgavas oppbygning	2
2	Bakgrunn.....	4
2.1	Lik tilgang til trykt tekst – en lovfestet rett	4
2.2	Studiesituasjonen for de lesehandikappede	5
2.3	Den digitale lydbokstandarden DAISY	5
2.4	Behovet for en talesyntese på studentens vilkår	6
3	Teori.....	8
3.1	Språkteknologi i Norge.....	8
3.2	Norske talesynteser.....	9
3.3	Nærmere om talesyntese, særlig Filibuster.....	10
3.4	Nordisk evalueringsforskning på talesyntese	15
3.5	Evalueringsmetoder av stemme- og talekvalitet	16
3.5.1	<i>På segmentnivå</i>	<i>18</i>
3.5.2	<i>På setningsnivå</i>	<i>21</i>
3.5.3	<i>Prosodievaluering</i>	<i>24</i>
3.5.4	<i>Forståelsestester.....</i>	<i>25</i>
3.5.5	<i>Generelle evalueringsmetoder</i>	<i>25</i>
4	Metode for å undersøke talekvalitet	27
4.1	Min egen posisjon til forskningsfeltet	27
4.2	Valg av evalueringsmetode	27
4.3	Forskningsetiske vurderinger	28
4.4	Brukerundersøkelse	29
4.4.1	<i>Utforming av spørreskjema.....</i>	<i>29</i>
4.4.2	<i>Kartlegging av respondenter og datamaterialets størrelse.....</i>	<i>30</i>
4.4.3	<i>Beskrivelse av respondentene.....</i>	<i>32</i>
4.5	SUS-test.....	33
4.5.1	<i>Om testen.....</i>	<i>33</i>
4.5.2	<i>Respondentene.....</i>	<i>34</i>

4.5.3	<i>Testmaterialet</i>	35
4.5.4	<i>Gjennomføring</i>	41
4.5.5	<i>Utrekning av resultatene</i>	41
5	Resultat og diskusjon	43
5.1.1	<i>Brukerundersøkelse</i>	43
5.1.2	<i>SUS-test</i>	86
6	Oppsummering og konklusjoner	113
6.1	Tydelighet.....	113
6.2	Naturlighet	114
6.3	Behagelighet	115
6.4	Aksept.....	116
6.5	Velegnethet.....	116
7	Veien videre	118
8	Svakheter ved studien	121
	Tabelliste	122
	Figurliste	125
	Litteraturliste	127
	Vedlegg	141
1	Vedlegg til brukerundersøkelsen	141
1.1	Spørreskjema	141
2	Vedlegg til SUS-test	154
2.1	Intransitive setninger	154
2.2	Transitive setninger	154
2.3	Imperative setninger	154
2.4	Spørresetninger	155
2.5	Relative setninger	155
2.6	Setningsrekkefølgen testen ble gjennomført med	155

1 Innledning

1.1 Tilrettelagt litteratur og talesyntese

Personer med ei syns- eller lesehemning har gjennom norsk lov krav på tilgang til tilrettelagt litteratur. Ved å tilrettelegge litteratur tilgjengeliggjøres informasjon. Dette er ei forutsetning for samfunnsdeltakelse for mennesker med nedsatt funksjonsevne (NOU 2005:8, s. 21-22). Norsk lyd- og blindeskriftbibliotek (NLB) er et statlig bibliotek, underlagt Kulturdepartementet. NLB tilrettelegger og formidler folkebiblioteklitteratur og studielitteratur på høyskole- og universitetsnivå. Hovedsakelig tilrettelegger NLB litteratur som lydbøker, men format som punktskrift og elektronisk tekst benyttes også. NLBs lydbøker skiller seg fra kommersielle lydbøker ved at de tilrettelegges i den digitale lydbokstandarden DAISY. DAISY (*Digital Accessible Information System*) gjør det mulig å synkronisere tekst og lyd slik at bøker blir mer tilgjengelige for personer med et lesehandikap (Kawamura, 2006; Leith, 2006). I denne studien vil disse lydbøkene bli referert til som DAISY-bøker. NLB produserer DAISY-bøker enten ved å lese dem inn i lydstudio, eller ved hjelp av talesyntese.

Talesyntese benyttes kun til produksjon av tidsskrifter, aviser og studielitteratur. Over halvparten av studielitteraturen tilrettelegges med syntetisk tale. I 2008 inngikk NLB et samarbeid med det svenske Talboks- og punktskriftsbiblioteket (TPB) om å tilpasse deres ”tekst-til-tale”-system Filibuster fra svensk til norsk bokmål. Filibuster ble særlig utviklet for å kunne tolke og behandle tekstuelle utfordringer, som særpreger studielitteratur. Tidligere benyttet NLB den kommersielle talesyntesen til Acapela Group, Kari, i sin produksjon (Acapela Group, 2004). Høsten 2009 ble Kari erstattet med NLBs egen talesyntese, Brage.

I denne studien ble den brukeropplevde talekvaliteten til Brage undersøkt ved å gjennomføre en brukerundersøkelse med utgangspunkt i erfaringer og opplevelser syns- og lesehemmete studenter hadde hatt med denne talesyntesen. Talekvaliteten ble målt etter kriterier som er sentrale innenfor denne type evalueringsforskning: velegnethet, tydelighet, naturlighet, aksept og behagelighet. I tillegg til brukerundersøkelsen ble det gjennomført en tydelighetstest, der Brage ble sammenliknet med to andre norske talesynteser.

1.2 Denne studiens formål og forskningsspørsmål

Tilrettelegging av studielitteratur med syntetisk tale er ikke uproblematisk. Det er gjennomført flere studier som antyder at syntetisk tale kan gjøre teksten mindre tilgjengelig. Axmear (et al., 2005) undersøkte talepersepsjon av syntetisk og menneskelig tale hos engelskspråklige barn og voksne, og observerte at syntetisk tale ga en dårligere forståelse hos både barn og voksne. Det er også gjennomført studier som viser at avkoding av syntetisk tale krever større konsentrasjon enn menneskelig tale (Engstrand, 1993; Persson, 2004).

Formålet med denne studien var å få kunnskap om hvilke egenskaper som virket positivt og negativt inn på talekvaliteten, og hvordan studentene opplevde Brage som formidler av pensumlitteraturen deres. Siden talesynteser viser seg både å være vanskeligere å avkode og forringer talepersepsjonen sammenliknet med menneskelig tale, var det ønskelig å supplere brukerundersøkelsen med en tydelighetstest. To andre norske talesynteser ble benyttet som sammenlikningsgrunnlag. Ved å kartlegge styrker og svakheter ved systemet og stemmen, kunne man få viktig kunnskap om hvilke områder som burde følges opp i den videre utviklinga.

Forskningsspørsmålene for denne studien var:

- Hvordan opplever studenter med et lesehandikap talekvaliteten til Filibuster med Brage, vurdert etter sentrale kriterier innenfor evalueringsforskninga?
- Hvor tydelig er Brage sammenliknet med andre norske talesynteser?
- Hvor velegnet synes Brage å være til å formidle studielitteratur?
- Hvordan bør Filibuster med Brage videreutvikles?

1.3 Oppgavas oppbygning

Denne oppgava består av åtte kapitler. I kapittel 1 blir tema, formål og forskningsspørsmål presentert. I kapittel 2 vil bakgrunnen for studien bli gjort rede for, og i kapittel 3 vil relevant teori og tidligere forskning på feltet bli presentert.

I kapittel 4 vil metoden for denne studien bli beskrevet. Det vil her gjøres greie for min egen posisjon til forskningsfeltet, valg av metode og aktuelle forskningsetiske vurderinger for studien.

I kapittel 5 vil det empiriske materialet bli framstilt og diskutert. I kapittel 6 vil funnene fra studien bli oppsummert og trukket konklusjoner ut ifra.

I kapittel 7 vil det bli foreslått hvordan ”tekst-til-tale”-systemet Filibuster med den syntetiske stemmen Brage kan forbedres med utgangspunkt i funnene fra denne studien.

I kapittel 8 er det gjort rede for svakheter ved denne studien.

2 Bakgrunn

2.1 Lik tilgang til trykt tekst – en lovfestet rett

Ifølge åndsverklovas § 17 er syns- og lesehemmete studenter i Norge sikret samme tilgang til studielitteratur som studenter uten lesehandikap. Videre heter det i Universitets- og høyskolelova § 4-3 at utdanningsinstitusjonene har ansvar for at syns- og lesehemmete studenter får tilgang til studielitteratur i et format som kompenserer for funksjonsnedsettelsen deres. Studenter med nedsatt funksjonsevne har i tillegg et diskrimineringsvern etter den nye diskriminerings- og tilgjengelighetslova, som trådte i kraft 1. januar 2009.

Syns- og lesehemmete studenter har behov for tilpasset studielitteratur. NLB har hovedansvaret for å tilgjengeliggjøre og produsere studielitteratur på høyskole- og universitetsnivå. Biblioteket har en egen studenttjeneste som skal sikre studenter i høyere utdanning obligatorisk pensumlitteratur i et egnet format.

For å bli låner hos NLB må en ha et dokumentert lesehandikap. Et lesehandikap kan være forårsaket av synshemning, lærehemning eller fysisk funksjonshemning som hindrer tilgang til innholdet i trykt tekst (Barraud, 2010). Denne definisjonen omfatter blinde, svaksynte, dyslektiske eller personer med andre lesehandikap, som for eksempel cerebral parese, multippel sklerose, ADHD m.m. (Kunnskapsdepartementet, 2007). Lesehandikaket må være dokumentert av en fagperson. Dette kan være en lege, synspedagog, spesialpedagog eller andre. Blinde og svaksynte som er registrert i Norges Blindforbund behøver ikke slik dokumentasjon (Norsk lyd- og blindeskriftbibliotek, 2010a).

Ikke alle studentene har i dag rett til å bestille det obligatoriske pensumet sitt i et tilpasset format (Norsk lyd- og blindeskriftbibliotek, 2010b). Det skilles mellom studenter med produksjonsrett og studenter med lånerett. I kalenderåret 2009 var det registrert 392 studenter ved NLBs studenttjeneste (Engh og Ellefsen, 2010, s. 18). Av disse hadde 21 % produksjonsrett. Produksjonsretten er i dag forbeholdt studenter med en visusfunksjonen som er redusert til 6/18 på beste øye med beste korreksjon. Dette betyr at studenten enten må være blind eller ha vesentlig nedsatt synsevne for å ha krav på hele det obligatoriske pensumet sitt. Øvrige lesehandikap gir kun mulighet til å låne bøker som allerede er produsert. Blant de lesehandikappede studentene utgjør dyslektikere en betydelig del (Kristensen, Gjælde og

Knarlag, 2008, s. 6). Studenter med sterk dysleksi kan søke om utvidet lånerett. Dette vil gi tilsvarende rettigheter som studenter med produksjonsrett.

2.2 Studiesituasjonen for de lesehandikappede

I 2005 gjennomførte Brandt en undersøkelse om hvordan studenter med behov for tilrettelegging opplevde Kvalitetsreformen (Brandt, 2005). Enkelte av de syns- og lesehemmete studentene opplevde reformen som problematisk, særlig fordi de ikke fikk tilpasset studielitteratur i tide. Studentene opplevde ofte at pensumlistene ikke ble tilgjengeliggjort før ved studiestart, og i enkelte tilfeller at pensumlistene ble endret på kort varsel. Siden det kan være tidkrevende å tilrettelegge studielitteratur, vil slike faktorer føre til at studentene mottar pensumbøkene sine enda seinere enn studenter uten lesehandikap. Studien til Brandt belyser også problematikken rundt produksjonsrett kontra lånerett. Studenter uten produksjonsrett må låne eldre studie bøker, som blinde og svaksynte har bestilt fra tidligere studier (Brandt, 2005, s. 11-12). Studenter uten produksjonsrett vil dermed heller ikke kunne velge foretrukket tilretteleggingsformat.

2.3 Den digitale lydbokstandarden DAISY

For om lag 10 år siden gikk NLB over til digital produksjon og besluttet samtidig å tilrettelegge lydbøker med lydbokstandarden DAISY. Til forskjell fra vanlige lydbøker, er tekst og lyd synkronisert i ei DAISY-bok. Med DAISY-boka kan studenten lett forflytte seg mellom de ulike tekstnivåene, som kapitler, sider, noter m.m. (Leith, 2006). Ei DAISY-bok der hele tekstinholdet er synkronisert med lyd, kalles DAISY-fulltekstbok. I ei DAISY-fulltekstbok vil leseren kunne ha mulighet til å forflytte seg mellom alle tekstnivåene som er representert i den trykte boka. Ved å tilrettelegge studielitteratur i DAISY, vil syns- og lesehemmete få muligheten til å søke etter ord, sette inn bokmerker, få teksten uthevet enten på setnings- eller avsnittsnivå under avspilling, ved sida av ei rekke anna funksjonalitet.

NLB produserer DAISY-bøker enten med menneskelig eller syntetisk tale. Bøker som blir produsert med talesyntese, tilrettelegges som DAISY-fulltekstbøker. Denne

synkroniseringa gjøres automatisk. Derimot blir studielitteratur som leses inn av profesjonelle innlesere hos NLB i dag, av økonomiske og produksjonsmessige årsaker, kun synkronisert med tekst på ulike overskriftsnivå, ved paginering og noter. Skal hele tekstinnholdet i boka synkroniseres med lyd, må man i dag gjøre dette manuelt, noe som vil være svært tidkrevende og dyrt.

For at studenten skal få tilgang til all funksjonalitet ei DAISY-bok gir, må boka spilles av med en DAISY-spiller. En DAISY-spiller kan være stasjonær eller foreligge som mykvarer for enten PC og Mac. Nes (2007, s. 110-114) studerte nytteverdien av DAISY i 2007. I denne studien observerte hun at ikke alle DAISY-spillerne på markedet syntes å støtte all kjernefunksjonalitet i DAISY.

2.4 Behovet for en talesyntese på studentens vilkår

Produksjon med talesyntese er billigere og gir mulighet for å ferdigstille og tilgjengeliggjøre studielitteratur hurtigere enn dersom det samme pensumet skulle leses inn i lydstudio (Hals, 2008). For å imøtekomme de utfordringene som ble observert av Brandt (2005), implementerte NLB tilrettelegging av studielitteratur med talesyntese i 2004. Tidligere ble alle lydbøkene lest inn av profesjonelle innlesere med relevant faglig bakgrunn eller tilstrekkelig fagforståelse til å formidle innholdet på en akseptabel måte. Til sammenlikning ble 61 % av studielitteraturen i 2009 produsert med talesyntese (Engh og Ellefsen, 2010, s. 10). Studielitteratur utgjør en sentral del av bibliotekets produksjonsgrunnlag, og har høy produksjonsprioritet.

Men det finnes flere forskjellige talesynteser på det kommersielle markedet i dag, og disse kan være trenet oppimot forskjellige bruksområder. Talboks- og punktskriftsbiblioteket (TPB) i Sverige er en statlig institusjon, som tilbyr tjenester for svenske lesehandikappede tilsvarende tjenestene NLB tilbyr norske. I 2005 utviklet TPB sin egen talesyntese.

Bakgrunnen var behovet for en talesyntese som særlig kunne tolke og behandle tekstuelle utfordringer som særpreget studielitteratur. Biblioteket trengte et system hvor de selv hadde mulighet til å tilføye ny fagterminologi og sentrale navn innenfor hver enkelt studieretning etter behov. Resultatet ble "tekst-til-tale"-systemet *Filibuster*, som ble satt i produksjon med

den første svenske stemmen *Folke* i februar 2007 (Ericsson et al., 2007). I løpet av ett års tid hadde biblioteket produsert om lag 350 svenske studiebøker med Folke.

I 2008 ble NLB bevilget én million kroner av Kunnskapsdepartementet for å styrke tilbudet til syns- og lesehemmete studenter (Engh og Ellefsen, 2008, s. 3). NLB besluttet å bruke deler av disse midlene til et samarbeid med TPB for å tilpasse deres system til norsk bokmål (Engh og Ellefsen, 2009). Jeg ble allokert fra stillinga mi som lydbokprodusent ved NLB for å delta i dette samarbeidet (Tånnander og Sjölander, 2008).

Sammen med andre medarbeidere ved NLB har min deltakelse i dette prosjektet omfattet rekruttering av innleser til den norske stemmen (Olaussen og Haugen, 2008), ansvar for innspilling av taledatabase, samt tilpassing og utvidelse av systemets regelsett og fonetisk transkribering. Gjennom hele prosessen har NLB hatt et tett samarbeid med systemutviklerne, datalingvist Tånnander og taleteknolog Sjölander, ved TPB.

5. oktober 2009 ble en norsk versjon av ”tekst-til-tale”-systemet Filibuster med stemmen Brage lansert (Tånnander og Sjölander, 2008). NLB hadde endelig fått sin egen talesyntese tilpasset norske studenters behov.

3 Teori

3.1 Språkteknologi i Norge

Språkteknologi gjør det mulig å utvikle produkter og tjenester, som for eksempel automatiske systemer for diktering, taleproduksjon, korrektur, maskinoversettelse, og informasjonssøking (Nordgård, 1998). For personer med et funksjonshinder vil språkteknologiske produkter og tjenester gi muligheten for deltakelse i arbeidslivet og samfunnslivet for øvrig. Talesyntese er et slikt språkteknologisk produkt, som bistår synshemmete med å lese og skrive. Bruk av talesyntese har også vist seg å være til stor hjelp for andre med lese- og skrivevansker, blant anna dyslektikere (Fuglerud og Solheim, 2008).

For å kunne utvikle en talesyntese er det nødvendig med språkteknologiske ressurser som tekst- og talekorpora. Slike korpora er språksamlinger med informasjon som benyttes i denne sammenhengen for å kunne gjenskape karakteristiske trekk ved menneskelig tale (Johannessen, 2003, s. 141-158; Sperberg-McQueen og Burnard, 2004). Behovet for et eget nasjonalt tekstkorpus ble drøftet så tidlig som i begynnelsen av 80-tallet (Breivik, 2003, s. 1). Tekstkorpuset som ble lagt til grunn for å tilpasse ”tekst-til-tale”-systemet Filibuster fra svensk til norsk bokmål, var Norsk Språkbank (2010). Språkbanken administreres og videreutvikles i dag i regi av Nasjonalbiblioteket. Et slikt tekstkorpus var ei forutsetning for å kunne utvikle talesyntesen Brage (Tånnander og Sjölander, 2008).

Et norsk språkteknologisk fagmiljø kan spores tilbake til 60-tallet da forskere ved Universitetet i Bergen begynte å behandle språk datamaskinelt (Nordgård, 1998). I dag videreføres denne forskningen ved Universitetet. Tekstlaboratoriet ved Universitetet i Oslo er en annen sentral aktør innenfor norsk språkteknologisk forskning. Tekstlaboratoriet har utviklet språkteknologiske programvare og verktøy, og jobber særlig med databaser og korpora for talespråk, skriftspråk m.m. Ved sida av det akademiske fagmiljøet er det også ei rekke private aktører som utvikler språkteknologiske produkter og tjenester.

For ni år siden ble det opprettet et forskningsprogram for språkteknologi med navnet Kunnskapsutvikling for norsk språkteknologi (KUNSTI) (Maegaard et al., 2007). Formålet med KUNSTI-programmet var for det første å gjøre talt og skrevet norsk (og samisk) tilgjengelig for datamaskinell prosessering, og for det andre å styrke grunnforskningen innenfor

språkteknologiske fagfelt, som datalingvistikk, taleteknologi, informatikk, fonetikk og formell lingvistikk (Ahrenberg et al., 2001). Et sentralt forskningsprosjekt som ble gjennomført innenfor KUNSTI-programmet, var FONEMA-prosjektet (FONEMA, 2010). Formålet med dette prosjektet var å komme fram til et rammeverk for hvordan man skulle lage en mest mulig naturlig talesyntese for norsk. Dette resulterte blant annet i en lingvistisk modell for norsk prosodi i syntetisk tale m.m. (Svendsen et al., 2005).

Det siste tiåret er det blitt gjennomført en mengde språkteknologisk forskning i Norge, som for eksempel på tekst- og talekorpora m.m. (Nøklestad og Søfteland, 2007; Siniscalchi, Svendsen og Lee, 2008; Amdal et al., 2008; Mertens et al., 2009; Johannessen et al., 2009; Sikveland et al., 2010; Guevara, 2010).

3.2 Norske talesynteser

Det finnes i dag ei rekke forskjellige talesynteser for norsk. Syns- og lesehemmete kan velge mellom flere ulike syntetiske stemmer etter egne preferanser om kjønn, lesehastighet, stemmeleie m.m. I tabell 1 er de sentrale talesyntesene for norsk presentert.

Tabell 1. Oversikt over sentrale norske talesynteser.

TALESYNTSE	TEKNOLOGI	LEVERANDØR	KOMMENTAR
Kari	Skjøtesyntese med enhetsutvalgelse	Acapela Group	NLB produserte studielitteratur med Kari siden den ble lansert i 2004
Trygve	Difonsyntese	Babel Infovox	
Vegard	Difonsyntese	Babel Infovox	
Olav		Acapela Group	
Nora	Skjøtesyntese med enhetsutvalgelse	Nuance RealSpeak	Opprinnelig utviklet av det norske selskapet Nordisk Språkteknologi (NST), som nå er konkurs
Stine	Skjøtesyntese med enhetsutvalgelse	Nuance RealSpeak	
Henrik	Skjøtesyntese med enhetsutvalgelse (med frasespleising)	IBM	Opprinnelig utviklet av Nordisk Språkteknologi
Telenor Talsmann	Difonsyntese	Telenor	

Lingspeak Arne	Difonsyntese	Lingit	Samme stemme som Telenor Talsmann, men med forbedret tekstforprocessor. Følger med alle norske eksemplarer av Windows XP og Vista
Isak HTS		Lingit	
Henrik	Skjøtesyntese med enhetsutvalgelse (med ekspressivitet)	Loquendo	
Vilde	Skjøtesyntese med enhetsutvalgelse (med ekspressivitet)	Loquendo	
Taleboka TTS Mona	Skjøtesyntese	NorMedia AS	
eSpeak norsk	Formantvokoder	Utvikles av frivillige bidragsytere verden over	Opprinnelig kjent som "speak". Det norske bidraget er fremdeles ikke testet og godkjent (november, 2010)

Talesyntesen Henrik (IBM), Henrik (Loquendo) og Vilde (Loquendo) synes å representere morgendagens talesynteser. IBMs Henrik er supplert med frasespleising. Dette innebærer at syntetisk tale genereres ved å skjøte sammen hele fraser i stedet for mindre lydenheter, slik tilfellet er med andre skjøtesynteser (Kvale, 2006, s.22). Italienske Loquendo har forsket på hvordan en talesyntese kan gjøres mer levende og personlig. Loquendo hevder å være blant de første på det kommersielle markedet som nå tilbyr talesynteser med ekspressivitet. Loquendos norske stemmer Henrik og Vilde kan blant anna synge, hoste og le (Loquendo, 2010).

Sentrale begrep som skjøtesyntese, enhetsutvalgelse og lydenheter er hittil benyttet uten noen begrepsavklaring. Disse begrepene vil i det følgende bli beskrevet. Det vil her bli gjort rede for de ulike bestanddelene til et "tekst-til-tale"-system og deres egenskaper ved å trekke inn eksempler fra filibustersystemet.

3.3 Nærmere om talesyntese, særlig Filibuster

Syntetisk tale er kunstig gjengivelse av menneskelig tale. Talesyntese omtales også som "tekst-til-tale" (eller TTS, av eng. *text-to-speech*) (Taylor, 2009). Et datamaskinelt "tekst-til-tale"-system kan implementeres enten i mykvare eller hardvare. Et slikt system er avhengig av at teksten er elektronisk før den kan behandles. Trykt tekst kan konverteres elektronisk ved skanning. Når den elektroniske teksten "mates" gjennom systemet, vil alle tekstelementer som

ord, bokstaver, sifre, tegn og anna, bli tolket og konvertert til syntetisk tale. Det skilles mellom tre grunnleggende måter å lage en talesyntese på:

- formantsyntese,
- artikulatorkisk syntese og
- datadrevet skjøtesyntese (også kalt bølgeformssyntese og konkateneringssyntese)

I denne studien er det ikke nødvendig å gå nærmere inn på formantsyntese eller artikulatorkisk syntese. Disse er nærmere beskrevet hos Taylor (2009, s. 388-399, og 405-407). Filibuster ”tekst-til-tale”-system er en datadrevet skjøtesyntese med enhetsutvelgelse. Enhetsutvelgelse er en av flere måter å lage skjøtesyntese på. I en skjøtesyntese med enhetsutvelgelse genereres syntetisk tale ved at små lydenheter, som plukkes ut fra en taledatabase, konkateneres (av lat. *concatenare*, å kjede sammen, å knytte sammen). En taledatabase vil bestå av store mengder med taledata, for eksempel som sekvenser med setninger, fraser eller mindre syntaktiske enheter. I en taledatabase til en skjøtesyntese med enhetsutvelgelse, er alle talesekvenser fonemisk annotert. Fonemisk annotering innebærer å segmentere talesekvensene i mindre lydenheter. Disse lydenhetene kan enten segmenteres som fonem, difon, fonemhalvdel, morfem, ord, frase eller som ei hel setning (Amdal, Johnsen og Svendsen, 2006, s. 2; Endresen, Simonsen og Sveen, 1996, s. 96; Kristoffersen, Simonsen og Sveen, 2005, s. 206-208; Faarlund, 2005, s. 63-71).

Taledatabasen til Filibuster utgjør nesten 570.000 slike lydsegmenter. Den fonemiske annoteringa av taledatabasen gjøres vanligvis automatisk ved hjelp av en talegjenkjenner. En slik metode vil kreve manuell korrigering i etterkant (Amdal, 2005, s. 13). Dette er nærmere beskrevet hos Black (2002).

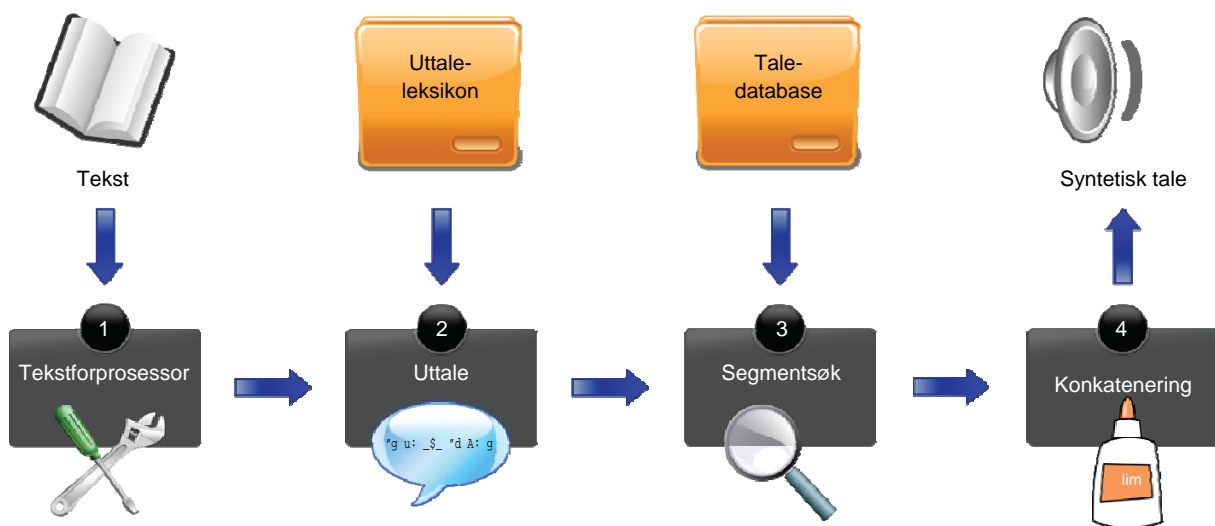
Oppretting av en taledatabase er omfattende. For å kunne gjenskape naturlig tale i forskjellige setningskonstruksjoner, bør man under innspillinga av en taledatabase sørge for å opprette et manuskript som sikrer at alle tenkelige prosodiske variasjoner for hver eneste lyd blir representert. Hvordan man ønsker å samle inn taledata er opp til en selv. Dersom man ønsker at ei fortellende setning skal leses med andre prosodiske egenskaper enn ved ei imperativsetning, er dette fullt mulig (Kristoffersen, Simonsen og Sveen, 2005, s. 106-109). Men desto flere variasjoner man ønsker, desto mer omfattende blir innsamlinga av taledata. Da taledatabasen er ferdig innspilt vil all informasjon om prosodisk variasjon og anna være

lagret sammen med hvert lydsegment i taledatabasen. En taledatabase vil dermed vanligvis bestå av mange timer med innlest tale. Filibuster består av over 26 timer innlest tale.

Det er flere utfordringer knyttet til å lage en naturlig skjøtesyntese. En skjøtesyntese genererer kunstig tale ved å lime sammen små lydenheter fra en database med innleste sekvenser med menneskelig tale. Talesekvensene vil med stor sannsynlighet være lest med uønskete varierende prosodi. Dette må forstås ut ifra at det vil være svært vanskelig, om ikke umulig å klare å lese setning etter setning i eksakt samme toneleie, lesehastighet, rytme, etc. Slik uønsket variasjon vil kunne resultere i en unaturlig, og dermed også utydelig talesyntese, når tale genereres ved å lime sammen mange lydenheter med varierende kvalitet. Ei anna utfordring kan være ”skjøtestøy”. Dette er nærmere beskrevet hos Kvale (2006).

Figur 1 nedenfor hvis hvordan en tekst konverteres til kunstig tale ved å bli tolket og prosessert av et ”tekst-til-tale”-system.

Figur 1. Figuren viser veien fra tekst til tale. Etter modell opprinnelig laget av Tånnander og Sjölander ved TPB.



Teksten analyseres og behandles først av (1) en tekstforprosessor. Forprosessorens hovedoppgave er å definere forskjellige tekstuelle elementer, som for eksempel akronymer, romertall, forkortelser, tall og talluttrykk, homografer, innslag med fremmedspråk, skrivefeil, osv. (Taylor, 2009, s. 95-105; Allen, Hunnicutt og Klatt, 1987, s. 55-57). Denne utfordrende

prosessen kalles *tekstnormalisering*, eller *tekstforprosessering*. I denne prosessen er det mye som kan gå galt. Se for eksempel på disse fremmedspråklige egennavnene: *Leicester*, *Edinburgh* og *Begin*. Selv for et menneske kan det være vanskelig å uttale egennavn korrekt dersom man ikke vet hvordan de skal uttales. Man kan ikke alltid gjette seg fram til riktig uttale bare ved å se på ortografien i enkelte tilfeller. Et annet problem med egennavn er at bærere av samme familienavn kan uttale det ulikt. For eksempel familienavnet mitt, *Olaussen*, vil jeg selv uttale med en diftong, *au*, mens andre kanskje vil gi det en monoftongisert uttale.

Etter at tekstforprosessoren har analysert og behandlet alle slike elementer, skal det deretter (2) bestemmes riktig uttale til hvert ord, tegn og andre elementer i teksten ved å hente uttaleinformasjon fra uttaleleksikonet. I uttaleleksikonet til Filibuster er informasjonen om hvordan et ord skal uttales, transkribert i det fonetiske alfabetet SAMPA (Wells, 2005). SAMPA er et akronym og står for *Speech Assessment Methods Phonetic Alphabet*. Dette fonetiske alfabetet er tilpasset datamaskinell lesing. Dessuten er det et praktisk alfabet, sammenliknet med for eksempel IPA (International Phonetic Alphabet) eller andre fonetiske alfabet, ettersom man har direkte tilgang til alle lydtegnene via et standard QWERTY-tastatur.

I drøftinga av resultatene fra denne studien er uttale gjengitt fonetisk med SAMPA. En full oversikt over norske SAMPA-symboler finnes på nettsidene til Department of Speech, Hearing and Phonetic Sciences ved University College London (Wells, 2005). Nedenfor er det gitt et eksempel på hvordan setninga *Nå kommer vinteren* ville ha sett ut i SAMPA. Ordgrensemarkøren # brukes vanligvis i SAMPA, men er i Filibuster erstattet med `_$` siden # har en annen funksjon i programmeringsspråket som tekstforprosessoren er distribuert i.

"n o: _\$ "k O \$ m @ r _\$ "v I n \$ t @ \$ r @ n

Uttaleleksikon omtales også som en leksikalsk database. Den leksikalske databasen til Skjøtesyntesen Filibuster er delt inn i følgende underleksikon:

- **Hovedleksikon**
Inneholder norske bokmålsord, og visse nynorske ord.
- **Navnleksikon**
Inneholder norske egennavn, så vel som engelske, tyske, franske og andre fremmedspråklige egennavn.
- **Fremmedspråklig leksikon**

Norsk Filibuster har i dag kun ett fremmedspråklig leksikon – engelsk.

- **Flerordsleksikon**

Navnet samsvarer ikke helt med innholdet og bruken av dette. I tillegg til høyfrekvente flerordsuttrykk i norsk bokmål, brukes dette leksikonet også til å sørge for at spesielle egennavn som *Napoleon III* og spesielle tekstelementer som [...] prosesseres korrekt.

- **Ekstraleksikon**

Provisorisk leksikon, som for eksempel kan brukes til ordforekomster som er unike for ei bok. Aktuelt å bruke dersom Brage skal lese forkortelsen *g.* som *gram* i ei kokebok. Men det vil være uheldig å unngå å fjerne dette dersom ei oppgavesamling i matematikk skal produseres.

I denne studien brukes gjerne begrepet *uttaleleksikon* synonymt med *hovedleksikon*. Ord som ikke er oppført i uttaleleksikonet annoteres med fonologisk informasjon fra en automatisk uttalegenerator. Dette gjøres på grunnlag av fonologisk og morfologisk analyse av teksten. Tekstinnholdet vil bli delt inn i syntaktiske enheter som ulike fraser, innskutte parenteser, leddsetninger og hovedsetninger (Kristoffersen, Simonsen og Sveen, 2005, s. 309-365). Denne prosessen kalles *tekst-til-fonem-* eller *grafem-til-fonem-konvertering*. Hver setning annoteres også med prosodisk uttaleinformasjon, for eksempel dersom setninga er ei spørresetning eller ei fortellende setning. Kvale (2006, s. 15) hevder at en intelligent tekstforprosessor trenger nærmere tusen slike uttaleregler for å håndtere det norske språket. Slike uttaleregler lages ofte ved å bruke den morfologiske informasjonen som er lagret sammen med den ortografiske og fonetiske informasjonen i uttaleleksikonet. Den morfologiske informasjonen i uttaleleksikonet til Filibuster er i dag annotert med taggsettet fra Stockholm Umeå-Korpuset (SUC versjon 2.0). SUC-taggsettet brukes til å ordklassifisere alle oppføringene i uttaleleksikonet, og å gi andre nødvendige opplysninger, som for eksempel om oppføringenes bøyningsform (Forsbom og Wilhelmsson, 2008, s. 39-40). Nedenfor er det vist hvordan de tre første kolonnene til ordene *nå*, *kommer* og *vinteren*, ser ut i uttaleleksikonet til Filibuster. I første kolonne angis hvordan det aktuelle ordet er skrevet (ortografisk informasjon). I andre kolonne er ordets lydstruktur transkribert i SAMPA, og angir hvordan ordet skal uttales (fonologisk informasjon). I tredje kolonne angir SUC-taggene hvilken ordklasse ordet tilhører, ordets bøyningsformer m.m. (morfologisk informasjon). Uttaleleksikonet til Filibuster består av mange flere kolonner, men disse vil ikke bli beskrevet ettersom de ikke er relevante for denne studien. Kolonnene er atskilt med en tabulator. Legg merke til at substantivet *vinteren* er morfologisk annotert som utrum. Dette vil bli nærmere forklart i forbindelse med beskrivelse av testmaterialet til SUS-testen, jf. Kapittel 4.5.3, s. 39. Mer informasjon om SUC-taggsettet finnes hos Ejerhed og Källgren (2006).

nå "n o: AB
kommer "k O \$ m @ r VB PRS AKT
vinteren "v I n \$ t @ \$ r @ n NN UTR SIN DEF NOM

Tekstforprosesseringa og grafem-til-fonem-konverteringa utgjør et ”tekst-til-tale”-systems første hoveddel. Neste hoveddel består i å konvertere all denne uttaleinformasjon om til syntetisk lyd, som beskrevet tidligere. Systemet (3) plukker ut egnete lydsegment, og (4) limer disse segmentene sammen. Disse konkatenererte segmentene vil gi en sammenhengende kunstig tale (van Santen et al., 1997, s. 261-262).

3.4 Nordisk evalueringsforskning på talesyntese

I Norge er det gjort ei rekke studier særlig med tilnærming til hvordan personer med et lesehandikap kan ved hjelp av taleproduksjon delta i arbeidslivet og informasjonssamfunnet, hvordan talesyntese vil være et nyttig hjelpemiddel i forbindelse med lese- og skriveopplæring, og hvordan talesyntese kan bidra til å imøtekomme noen av utfordringene lesehemmete står overfor i forhold til bruk av IKT, teletjenester og liknende (Årsand et al., 2004; Karlsen, 2007; Hansen, 2008; Fuglerud og Solheim, 2008). Det er også forsket en del på hvordan man bør utvikle tekst- og talekorpora, samt hvordan man kan forbedre talekvaliteten til en talesyntese (Amdal og Svendsen, 2006; Amdal et al., 2008; Svendsen et al., 2005; Meen, Svendsen og Natvig, 2005; Amdal, Johnsen og Svendsen, 2006).

Men det synes ikke å være utført noen studier i Norge hvor formålet er å evaluere talekvaliteten til syntetiske stemmer med metoder, som er foreslått innen denne typen evalueringsforskning. Så vidt meg bekjent foreligger det heller ingen norske studier innenfor lingvistiske og nordiske studier som omhandler evaluering av talekvalitet til talesynteser.

Riktignok har IT-selskapet MediaLT gjennomført et prosjekt hvor bruk av syntetisk tale i DAISY- fulltekstbøker ble undersøkt (Tollefsen 2006). Formålet med prosjektet var blant annet å utvikle og evaluere metoder og standarder for produksjon av DAISY-

fulltekstbøker. Nes (2007) har som nevnt også studert DAISY-standarden. Hun evaluerte den funksjonelle nytteverdien til DAISY-bøker for elever i det norske skolesystemet.

I Norden for øvrig derimot er det forsket mer på å evaluere talekvalitet med metoder fra evalueringsforskninga. I Finland gransket Lemmetty (1999) talesynteser i masteroppgava si ved Det tekniske universitetet i Helsinki. Han gir blant anna en god oversikt over de mest anvendte evalueringsmetodene for å måle talesynteser. I Sverige studerte Ståhl (2009) i sin magisteroppsats i allmenn språkvitenskap hvordan voksne synshemmete forsto tekster som ble produsert med syntetisk tale, sammenliknet med menneskelig tale. Ståhl målte blant anna tydeligheten og naturligheten i den svenske stemmen *Folke*, som er en del av Filibuster, og sammenliknet med stemmen til ”mannen bak Folke”. Han studerte også sammenhengen mellom talepersepsjon og konsentrasjon. Persson (2004) målte talekvaliteten til fire svenske talesynteser i masteroppgava si i datalingvistik. I hennes studie av talesyntese til bruk i lydbøker, målte hun flere sider ved talekvaliteten, som for eksempel anstrengelse, aksept og preferanser, lesehastighet og egenskaper ved stemmekvaliteten. Det er utført ei rekke studier i Sverige, hvor man enten har foreslått nye metoder for å evaluere talekvalitet eller hvor man har brukt gjeldende metoder for å måle ulike sider ved talekvaliteten til en talesyntese (Carlson, Granström og Nord, 1990; Carlson et al., 1992; Neovius og Raghavendra, 1993; Engstrand, 1993; Beskow et al. 1997).

En annen studie som må presenteres er brukerundersøkelsen som Cryer og Home (2009) gjennomførte for å kartlegge britiske lesehemmete lånere sine holdninger til syntetisk opplesing. Den er riktignok ikke utført i Norden, men allikevel svært sentral som sammenlikningsgrunnlag for brukerundersøkelsen i denne studien. De gjennomførte en generell undersøkelse for å måle ulike sider ved talekvaliteten, særlig oppimot evalueringskriteriet om aksept. Cryer og Home har arbeidet mye med evaluering av talesynteser i forskninga si ved det britiske ”søsterbiblioteket” til NLB, The Royal National Institute of Blind People (RNIB) (Cryer og Home, 2008; Cryer og Home, 2010).

3.5 Evaluering av stemme- og talekvalitet

Innenfor evalueringsforskning på talesyntese refereres det gjerne til *voice and speech quality* og *speech perception* (Jekosch, 2005, s.1-9). I denne studien er først og fremst talekvalitet og

talepersepsjon sentrale. Forskninga av stemme- og talekvalitet og talepersepsjon er ifølge Jekosch et relativt ungt forskningsfelt med røtter innenfor psykoakustikk, metrologi og lingvistikk. Mens stemmekvalitet omfatter alle karakteristiske egenskaper og særpreg som vil definere kvaliteten til en syntetisk stemme, må talekvalitet defineres som et resultat av disse egenskapene og særpregene. Ut ifra dette mener Jekosch (2005, s. 5-9) at talekvalitet må måles på grunnlag av hvor velegnet en syntetisk stemme er til å innfri de kravene og forventningene som stilles til den. Talekvalitet er et komplekst og sammensatt begrep, hvor flere faktorer vil påvirke hvor velegnet en talesyntese synes å være. Nøkkelordet i denne definisjonen er *velegnethet*. Kriteriet om velegnethet må anses som et overordnet evalueringskriterium (Mariniak, 1993). Og ettersom talesynteser har forskjellig bruksområde, vil velegnethetskriteriet variere fra talesyntese til talesyntese (King, 2007, s. 130-131).

Det er gjennomført omfattende mengder forskning på evalueringsmetodikk for å måle talekvaliteten av talesynteser (Campbell, 2007, s. 45). Det er opprettet flere egne fora for å organisere arbeidsgrupper og seminarer hvor systemutviklere og forskere over hele verden samles for å samarbeide om å forbedre gjeldende evalueringsmetoder eller foreslå nye (Campbell, 2007, s. 43-45; King, 2007, s. 126-128). Tre sentrale arrangører er *The Blizzard Challenge*, *COCOSDA* (<http://www.cocosda.org>), og *ISCA* (<http://www.isca-speech.org>). Ei rekke andre arrangører nevnes i Campbell (2007).

En talesyntese kan evalueres enten diagnostisk eller komparativt, subjektivt eller objektivt, på komponentnivå eller generelt. Dette kan gjøres via Internett eller ”ansikt-til-ansikt”, med hodetelefoner eller høyttalere, med de faktiske brukerne eller innleide lyttere, spesialister eller ikke-spesialister, gruppevis eller enkeltvis. Dessuten kan selve målinga gjøres på ulike nivå, som fonem-, ord- eller setningsnivå. Det er utarbeidet ei rekke forskjellige evalueringsmetoder med ulik hensikt og fokus. Blant de mest utbredte er ulike typer av subjektive lyttetester (Taylor, 2009).

Det synes å være en felles enighet blant forskerne om at grunnlaget for evaluering av en talesyntese bør være dens evne til å generere syntetisk tale som er mest mulig tilnærmet menneskelig tale (van Santen et al., 1997, s. 554-557). Dette refereres til som kriteriet om *naturlighet*. Eldre talesynteser ble hovedsakelig målt ut ifra hvor tydelige de var, framfor hvor naturlige de var (Jekosch, 2005). Det bør her nevnes at Mattingly (sitert i Campbell, 2007, s. 36) så tidlig som i 1974 gjorde et poeng av hvor viktig en naturlig talekvalitet er når en menneskestemme skal gjenskapes syntetisk:

From our point of view it is not physical realism but psychological acceptability which is the proper evidence for correctness at the phonological and phonetic levels, just as it is on the syntactic level.

Men en naturlig gjengivelse av menneskestemmen vil være avhengig av at den også er tydelig. Evalueringskriteriet om *tydelighet* er nært knyttet til naturlighet. Kriteriet om tydelighet defineres av Bosshart (sitert i Jekosch, 2005, s. 100) som:

A statement is said to be comprehensible when it can be reiterated by the recipient in the sense of the sender on a certain level of comprehension.

I tillegg til kriteriene om velegnethet, naturlighet og tydelighet, foreslår Campbell et annet sentralt evalueringskriterium: *likeability* (Campbell, 2007, s. 36). For studenter som må lytte time etter time med en syntetisk stemme de ikke liker, kan i verste fall være uutholdelig.

Men selv om disse evalueringskriteriene er benyttet i stor utstrekning, er de bare forslag. Fremdeles foreligger det ingen enighet om hvilke evalueringskriterier som skal legges til grunn, og dessuten hvilke krav en skal stille til testmaterialet (Black og Lenzo, 2010). Ei rekke forskjellige evalueringskriterier er forslått, men som King (2007) poengterte, må man selv avgjøre hvilke kriterier som synes å være mest egnet.

I det følgende vil det redegjøres for noen av de sentrale evalueringsmetodene. Det er svært mange forskjellige tester å velge mellom, avhengig av hva man vil måle. Man skiller gjerne mellom evaluering på segmentnivå eller på setningsnivå, evaluering av prosodiske egenskaper, forståelsestester og generelle tester som måler den helhetlige talekvaliteten. For mer informasjon om måling av talekvalitet og framstilling av andre evalueringsmetoder, vises det til Jekosch (2005, s. 89-103), Cryer og Home (2010), Dybkjær, Hemsén og Minker (2008), og Dynastat (2010).

3.5.1 På segmentnivå

Felles for disse testene er at kun tydeligheten av ett enkelt lydsegment blir målt. De mest utbredte testene på segmentnivå er rimtester og tester med ”tulleord”. *Rimtester* har flere fordeler. De er konsentrerte og effektive, krever liten tid å gjennomføre, og informanter uten noen form for trening kan delta. Dessuten vil man oppnå resultat med et relativt lite utvalg, dvs. vanligvis mellom 10 til 20 informanter. Læringseffekten kan også måles. Rimtester er

økonomisk rimelige å gjennomføre, og enkle å administrere. Ei ulempe er at siden det kun brukes enstavelsesord, kan bare enkeltkonsonanter bli testet. Rimtesten krever også at man gjennomfører testen etter et forhåndsdefinert og ”lukket” testmateriale. Et slikt testmateriale vil dermed kanskje ikke vise svakheter som er reelle for det faktiske bruksområdet til en talesyntese. Og siden testmaterialet er forhåndsdefinert vil informanter som har deltatt i en tidligere rimtest kanskje huske enkelte riktige svar (Jekosch, 1993, s. 1387-1394). For å unngå disse ulempene anbefaler Jekosch *gruppegjenkjenningstesten* (Jekosch, 1992, s. 205-208), som er beskrevet nedenfor. Rimtester er tilgjengelige på mange språk.

En mye brukt rimtest, er *den diagnostiske rimtesten*, som ble presentert av Fairbanks i 1958. Den diagnostiske rimtesten inkluderer et forhåndsdefinert sett med isolerte ord. Formålet med denne testen er å måle konsonantisk tydelighet i begynnelsen av ord (Goldstein, 1995, s. 225-244). Testsettet inneholder 96 ordpar som skiller seg fra hverandre med ett enkelt lydlig trekk; konsonantlyden i begynnelsen av hvert ord i et ordpar er forskjellig fra hverandre. Ordparene er konstruert for å måle kvaliteten av seks forskjellige fonetiske egenskaper ved en konsonantlyd: stemthet, nasalitet, kvantitet, visling, aksent og innsnevring (Kristoffersen, Simonsen og Sveen, 2005, s. 170-175; Endresen, Simonsen og Sveen, 1996, s. 293-297; Kulbrandstad, 1993, s. 31-36). Det spilles av ett og ett ord for informanten, som skal gjengi det ordet han tror han hørte. Resultatene vises enten prosentvis eller i en matrise, for å kunne observere hvilke konsonantlyder som særlig forveksles. Diagnostiske rimtester er nyttige under utviklinga av et system. Testen gir kun informasjon om konsonantlyden i opptakta til enstavelsesord. Den måler ikke konsonantlyder i midtre eller endelig stavelse, og forteller heller ingenting om kvaliteten av vokallyder eller prosodiske forhold. Av den grunn egner denne testen seg ikke som ei generell evaluering av hvor tydelig en talesyntese er på segmentnivå. Og ettersom testmaterialet er avgrenset vil man ikke kunne måle alle tenkelige konsonantiske forvekslingstilfeller (Carlson, Granström og Nord, 1990, s. 271-277).

En annen utbredt type rimtester som er mye brukt, er *de modifiserte rimtestene*. De modifiserte rimtestene likner de diagnostiske, men til forskjell måler de modifiserte både konsonanter i første og i siste stavelse av et ord (Goldstein, 1995). Disse testene tar i bruk et forhåndsdefinert testmateriale med 50 ordsett, som hvert består av seks enstavelsesord. Dette utgjør til sammen 300 ord. Informanten skal lytte til ett og ett sett om gangen og markere det ordet han tror han hørte i et *multiple choice*-svarark. Første halvdel av et sett har til hensikt å måle tydeligheten av konsonantene i første stavelse av et ord, og tilsvarende måles

konsonantene i siste stavelse av den siste halvdel av et sett. Resultatene framstilles på samme måte som ved de diagnostiske rimtestene (Shiga, Hara og Nitta, 1994, s. 1783-1786).

En annen rimtest er *den diagnostiske midtkonsonanttesten* (DMCT), som likner de diagnostiske rimtestene. Testmaterialet til en diagnostisk midtkonsonanttest består av 96 par med tostavelsesord. Disse plukkes ut på en slik måte at den intervokaliske konsonanten, dvs. den konsonanten som står mellom to vokaler, skiller parene fra hverandre. Den diagnostiske midtkonsonanttesten har, tilsvarende den diagnostiske rimtesten, også til formål å studere de samme seks ulike fonetiske egenskaper ved en konsonantlyd (Dynastat, 2010).

Testen som Jekosch anbefalte, kalles *gruppegjenkjenningstesten*, og ble utviklet under ESPRIT-prosjektet SAM (Jekosch, 1993). Testen har ei statistisk tilnærming. Testmaterialet er ikke forhåndsdefinert. Nytt testmaterialet opprettes for hver test. Det er utviklet en ordgenerator som konstruerer det nødvendige testmaterialet. Testansvarlig avgjør selv hvor mange ord testmaterialet skal bestå av, og hvilke stavelsesstrukturer som skal anvendes. De ulike stavelsesstrukturene følger mønstrene K_nV , K_nVK_n , VK_n , der testansvarlig selv angir antallet n (Jekosch, 2005, s. 98). På denne måten kan en selv bestemme størrelsene på konsonantgruppene, for eksempel ved å konstruere konsonantgrupper som ofte går igjen i et språk. Testmaterialet vil hovedsakelig bestå av meningsløse ord. Testutviklerne har også laget et program som automatisk regner ut feilrater og lister opp resultatene i en matrise (Jekosch, 1992).

En annen sentral test for å måle tydeligheten av lydsegment, er *SAM standard segmenttest*. Testmaterialet til denne testen omfatter enkeltord, hovedsakelig uten noen form for mening, som enten har strukturen KV , VKV eller VK . K står for konsonant og inkluderer alle mulige konsonantlyder. V står for vokal, og inkluderer bare vokallydene $/A/$, $/A:/$, $/I/$, $/i:/$, $/u0/$ og $/\}:/$, dvs. lange og korte a-, i- og u-vokallyder. Testmaterialet er tilgjengelig for engelsk, tysk, svensk og nederlandsk. Informanten skal lytte til enkeltordene og skrive ned det han tror han hørte. Besvarelsen gjengis åpen uten noen begrensninger. Det er kun tydeligheten av konsonantlydene som måles i denne testen, både de i begynnelsen, i midten og i slutten av ord (Jekosch, 2005, s. 97).

Ei fonetisk balansert ordliste er en test hvor testmaterialet omfatter forhåndsdefinerte enstavelsesord. Ei slik ordliste ble for første gang foreslått ved Harvard University i 40-årene. Denne ordlista er statistisk balansert ved at enstavelsesord, som alltid synes å mangle i

besvarelsene til informantene, eller som alltid blir korrekt gjengitt, fjernes fra ordlista. Testen gjennomføres tilsvarende de øvrige testene; ordene spilles av for informantene, som skal gjengi det de hørte. Det er ingen begrensninger i besvarelsene. Siden den første testen er det blitt utviklet flere balanserte ordlister, som for eksempel *Phonetically Balanced-50 word discrimination test*, *PB-50*, som består av femti enstavelsesord med tilsvarende frekvens som tidligere ordlister (Goldstein, 1995).

En test som er mye brukt er den såkalte *Nonsense words*-testen. Hensikten med å gjennomføre en test bestående av ”tulleord”, er å måle tydeligheten ved overgangen mellom vokaler og konsonanter. Denne testen gir høye feilrater og et godt diagnostisk materiale, særlig dersom informantene ikke gis noen restriksjoner i forhold til besvarelsen. Vanligvis brukes ei liste med strukturene VK, KV, VKV eller KVK. Det er mulig å plukke ut lengre ord til testmaterialet også, dersom det er ønskelig (Carlson, Granström og Nord, 1990).

Den siste testen som skal nevnes i forbindelse med å måle tydelighet på segmentnivå, er *Bellcore-testen*. Dette er en diagnostisk og komparativ test. Hensikten med testen er å måle tydeligheten av sekvenser med én eller flere konsonanter i første eller siste stavelse til et ord. Testen er så vidt meg bekjent bare tilgjengelig for engelsk. Testmaterialet består av et ordnet sett med 312 K_nVK_n -ordenheter, der n står for et bestemt antall. Ordenhetene har bare én stavelse og opptrer parvis på en slik måte at en av ordenhetene er meningsbærende og den andre ikke, som for eksempel *swan – swog*, *warmth – dorth*, *dropped – globbed*. Informantene skal gjengi det de hørte. Svarene er åpne uten noen restriksjoner. Formålet med denne testen er å måle hvor lett hver enkelt konsonantgruppe i enstavelsesord kan gjenkjennes (Jekosch, 2005, s. 98).

3.5.2 På setningsnivå

De neste testene som skal presenteres har alle til felles at de måler tydelighet på setningsnivå. Et særtrekk som ofte går igjen i slike tester, er at det skal lages setninger bestående av høyfrekvente ord. Ei særlig ulempe med slike tester, som ikke synes å være gjeldende for tester som måler tydelighet på segmentnivå, er at man kan få et datamateriale med korrekte svar selv om det er ordenheter som mangler. Dette gjelder hovedsakelig tester som tar i bruk meningsbærende setninger (Pisoni og Hunnicutt, 1980, s. 572-575).

Den første testen for måling av talekvalitet på setningsnivå, er *Harvard psykoakustiske setninger*. Denne testen består av et forhåndsdefinert og ”lukket” sett med hundre meningsbærende setninger. Setningene ble opprettet på en slik måte at kun de mest høyfrekvente enkeltlydene i det engelske språket skulle brukes. Noen eksempler fra setningssettet:

Rice is often served in round bowls. Glue the sheet at the dark blue background.

These days a chicken leg is a rare dish. It's easy to tell the depth of a well.

Hensikten med denne testen er å måle hvor tydelig hvert ord er på setningsnivå. Informanten skal skrive ned det de tror de hørte (Jekosch, 2005, s. 99). Fordelen med testen er at den er lett å gjennomføre, og krever ingen trening av informantene på forhånd. Enkelte forskere har kommentert at selv om setningssettet er omfattende, kan det allikevel tenkes at informantene kan være kjent med testmaterialet fra tidligere, ettersom setningene er forhåndsdefinerte (Allen, Hunnicutt og Klatt, 1987).

Den neste testen som skal presenteres er de anerkjente *Haskins-setningene*. Denne testen er også utviklet for å teste tydelighet på setnings- eller ordnivå. Tilsvarende Harvard psykoakustiske setninger består også Haskins-setningene av hundre setninger, som er laget ved å ta utgangspunkt i de mest høyfrekvente enkeltlydene i engelsk (Gibbon, Moore og Winski, 1997, s. 214). Men til forskjell fra Harvard psykoakustiske setninger er ikke Haskins-setningene meningsbærende. Lytterne skal skrive ned det de hørte (Jekosch, 2005, s. 99). Eksempler fra dette setningssettet:

The great car met the milk. The old corn cost the blood.

The old farm cost the blood. The short arm sent the cow.

Den tredje testen som skal presenteres er *de semantisk uforutsigbare setningene* (SUS) (Benoît, Grice og Hazan, 1995, s. 381-392). SUS-testen ble utviklet av SAM-gruppa, og først introdusert for de europeiske språkene i 1996. I 2005 ble den presentert for the Blizzard Challenge, og i 2006 for engelsk. Hensikten med semantisk uforutsigbare setninger er å fjerne all semantisk informasjon fra testmaterialet, slik at kun den syntaktiske gjenstår. Tanken er at slik informasjon kan svekke påliteligheten til funnene ettersom informanten får kontekstuell hjelp til å tyde setningene. Dersom ett ord er utydelig i ei meningsbærende setning, kan allikevel informantene gjette seg fram til riktige ordet (Frawley, 1992, s. 12-13). Feilraten er

vanligvis dermed høyere i tydelighetstester som tar i bruk setninger uten meningsinnhold, sammenliknet med tester som for eksempel Harvard psykoakustiske setninger.

Testmaterialet i SUS-testen består av setninger som opprettes etter fem fastsatte syntaktiske strukturer. Ordene som plukkes ut for å opprette disse setningene er hovedsakelig enstavelsesord, men to- eller flerstavelsesord kan brukes dersom det er nødvendig for testspråket. Tanken bak dette er å forhindre testen i å bli en hukommelsestest. SAM-gruppa oppfordrer til at testmaterialet bør bestå av minst 50-60 setninger. Ordene som plukkes ut skal være de mest høyfrekvente i språket som skal testes. Det stilles videre ei rekke restriksjoner til hvilke ord som skal brukes. Disse vil bli nærmere omtalt i metodekapitlet, jf. Kapittel 4.5.3, s. 35-41. Tanken med å bruke fem ulike setningsstrukturer er å unngå risikoen for at lytterne skal kunne gjette seg fram til et gjentakende mønster. Et gjentakende mønster vil gi lytterne syntaktisk informasjon, som også vil hjelpe dem til å gjette seg fram til utydelige ord. Dersom informantene gjenkjenner mønstret og vet at det må være et verbal som er utydelig, vil dette innvirke på testens pålitelighet. Testen gjennomføres på tilsvarende måte som mange andre tester, nettopp ved at informantene skal lytte til ei og ei setning og gjengi det de tror de hørte.

I tabell 2 nedenfor er de fem strukturene med eksempler, som Benoît, Grice og Hazan (1996) har foreslått, gjengitt. De har gitt eksempler både for engelsk, nederlandsk, fransk, tysk, italiensk og svensk. De engelske og svenske vil her bli framstilt.

Tabell 2. De fem syntaktiske strukturene i SUS-testen med engelske og svenske eksempler.

Syntaktisk struktur	Engelsk	Svensk
1. Intransitiv struktur	The table walked through the blue truth.	En stol dog till ett tomt hus.
2. Transitiv struktur	The strong way drank the day.	En klar bok sjøng en ko.
3. Imperativ struktur	Draw the house and the fact.	Se en bok och en sag.
4. Interrogativ struktur	How does the day love the bright word?	Hur blev en lukt ett snabbt hus?
5. Relativ struktur	The plane closed the fish that lived.	En plan at en fisk som sam.

3.5.3 Prosodievaluering

Bedømmelsen av prosodien til en talesyntese vil kunne si noe om hvor naturlig en syntetisk stemme høres ut. Men det kan være utfordrende å måle prosodiske forhold. De vil påvirkes av flere komponenter i et ”tekst-til-tale”-system, slik at det vil være vanskelig å foreslå gode metoder som vil kunne måle prosodien atskilt fra andre egenskaper og særpreg (Sonntag og Portele, 1997, s. 188-194). Ei enkel måling av prosodien kan for eksempel være å opprette et testmateriale bestående av utsagnssetninger, spørresetninger og imperativssetninger. Disse setningene kan spilles av for en informant som skal sondre mellom de ulike setningstypene (Lemmetty, 1999, s. 86).

I 1992 presenterte Grice m.fl. to tester for å måle prosodiske forhold: *SAM prosodisk formtest* og *SAM prosodisk funksjonstest* (Gibbon, Moore og Winski, 1997, s. 559-560). Formtesten gjør det mulig å foreta ei diagnostisk og komparativ måling av parameterne i et talesyntesystem, som definerer kvaliteten til bestemte konturer i setningsmelodien, uavhengig av den tekstuelle konteksten. Formtesten er designet for engelsk og italiensk. Testmaterialet består av meningsbærende ord, som varierte i forhold til vokallengde og stavelsesstruktur. Det gjeldende evalueringskriteriet for testen er kriteriet om naturlighet. Den prosodiske funksjonstesten er også ei diagnostisk og komparativ måling, slik som formtesten, men funksjonstesten er hovedsakelig beregnet for dialogsystemer, hvor talesyntese brukes i interaksjon mellom menneske og datamaskin. Hensikten med funksjonstesten er å kunne si noe om hvilke sider ved setningsmelodien som er velegnet i slik interaksjon. Testen er tilgjengelig for engelsk, men kan lett tilpasses andre språk.

En annen metode for å måle kvaliteten ved prosodiske trekk, er den flermodulære evalueringsmetoden som er beskrevet i EvaSy-prosjektet (Garcia et al., 2006). Dette er ei totalevaluering som inkluderer fire delmålinger: måling av grafem-til-fonem-konverteringa, måling av forståelighet, måling av prosodiske forhold, og til slutt ei helhetlig måling av den generelle talekvaliteten. Målinga av de prosodiske forholdene foretas ved å bruke et prosodisk overflyttingsparadigme, som innebærer at konturene i setningsmelodien til ulike setninger prøves på ei vanlig setning. Garcia m.fl. har demonstrert hvordan dette paradigmet fungerer ved å gjennomføre metoden på franske talesynteser.

3.5.4 Forståelsestester

Testmaterialet til forståelsestester består vanligvis enten av setninger eller tekstutdrag, som et avsnitt eller ei side med tekst. Dette blir spilt av for en informant med noen etterfølgende spørsmål til hva man nettopp hørte. Hensikten med forståelsestester er ikke å kartlegge hvilket enkeltfonem som er vanskeligere å gjenkjenne enn andre. Målinga foretas på grunnlag av hvor god evne en talesyntese har til å formidle en tekst (Lampert, 2004, s. 11-13).

3.5.5 Generelle evalueringsmetoder

Generelle evalueringsmetodene er utviklet for å gi ei helhetlig vurdering av talekvaliteten til en talesyntese. I slike metoder blir ikke målinga vurdert utelukkende ut ifra ett eller noen få evalueringskriterier, slik som tilfellet er med mange andre metoder.

Mean Opinion Score (MOS) er en metode som er mye brukt (Goldstein, 1995). Dette er en generell brukerundersøkelse, som er enkel å gjennomføre og gjør det lettere å si noe generelt om talekvaliteten til en talesyntese. MOS-testen gjennomføres ved at informantene skal gi vurdering til en lydsekvens. Denne vurderinga gjøres ut ifra en skala fra én til fem, hvor én er dårligst og fem er best. Denne skalaen er også kjent som ACR (*Absolute Category Rating*). Det er også gjort forsøk å bryte denne skalaen inn i mer beskrivende kategorier, som for eksempel *Veldig irriterende, irriterende, nokså irriterende, forståelig men ikke irriterende, og uforståelig*. En slik skala kalles DMOS (*Degradation MOS*) eller DCR (*Degradation Category*).

En annen måte å vurdere den generelle talekvaliteten på, er ved hjelp av *kategorisk estimering* (Kaur og Singh, 2010). Dette er en metode som forsøker å måle ei rekke forskjellige sider ved talen, uavhengig av hverandre. På samme måte som med MOS-metoden skal informantene gi vurderinga si med utgangspunkt i en skala som spenner seg fra *dårligst* til *best*. Kraft og Portele (sitert i Kaur og Singh, 2010) foreslår vurderingskategorier som *veldig lett, lett, vanskelig, veldig vanskelig* dersom det gjeldende evalueringskriteriet for vurderinga er tydelighet. Dersom man ønsker å vurdere uttale, lesehastighet, naturlighet, ordtrykk, tonelag, behagelighet eller annet, kan man tilpasse kategoriene slik at beskrivelsen passer til det man skal måle. Denne metoden er også enkel å gjennomføre og praktisk på den måten at den gir ei helhetlig individuell vurdering av talekvaliteten.

En annen metode for å vurdere talekvaliteten på generelt grunnlag, er metoden for omfangs- og forholdsestimering (referert til i engelsk faglitteratur som *Magnitude and Ratio Estimation*) (Pavlovic, Rossi og Espesser, 1990, s. 373-382). Resultatene beregnes og framstilles numerisk for å kunne vurdere omfanget av forskjellige sider ved lytteopplevelsen. Ifølge Pavlovic, Rossi og Espesser egner denne metoden seg like godt for å måle sanslige inntrykk, så vel som følelser knyttet til lytteopplevelsen. I motsetning til kategorisk estimering, vil forholdene som skal vurderes med omfangs- og forholdsestimering ikke være avgrenset med et forhåndsdefinert antall verdier og en "låst" skala. I denne metoden kan informantene selv velge fritt hvilken verdi de ønsker å gi et evalueringsobjekt (van Heuven og Pols, 1993, s. 356-357).

En annen metode med en kategorisk tilnærming til å måle talekvalitet generelt, er *ITU-T-metoden* (Gibbon, Moore og Winski, 1997, s. 248-249). Dette er en komparativ evalueringsmetode som forsøker å gi ei samlet vurdering av alt fra kategorier som for eksempel holdninger, aksept, helhetsinntrykk, konsentrasjon, forståelighet, artikulering, uttale, lesehastighet og behagelighet. Metoden foreslår også mulige spørsmål med egnete svaralternativ (Jekosch, 2005, s. 101-102).

4 Metode for å undersøke talekvalitet

4.1 Min egen posisjon til forskningsfeltet

Jeg har vært aktivt deltakende på flere områder under tilpassinga av Filibuster fra svensk til norsk bokmål, og har jobbet med systemet siden august 2008. I dag har jeg hovedansvaret for videreutviklinga av talesyntesen. Det vil være både fordeler og ulemper med ei slik tilknytning til evalueringsobjektet (Robson, 1993, s. 382). Fordelen med å gjennomføre ei evaluering av en talesyntese man selv har arbeidet med, var derimot at jeg hadde inngående kjennskap til systemet og områder som kunne innvirke på talekvaliteten. Men det innebar også ei utfordring i å ivareta ei objektiv og kritisk tilnærming til et evalueringsobjekt jeg selv har arbeidet med (Robson, 1993, s. 6-7).

4.2 Valg av evalueringsmetode

Valget av evalueringsmetode må foretas ut ifra hva som var ønskelig å kartlegge (King, 2007, s. 131-134). Siden Filibuster ble utviklet for å produsere tilrettelagt studielitteratur, var det naturlig å gjennomføre en brukerundersøkelse med utgangspunkt i studentenes erfaringer (King, 2007, s. 128-129). Dette skulle gi et datagrunnlag for å bedømme studentenes opplevelse av den generelle talekvaliteten. Gjennom en brukerundersøkelse ville sentrale evalueringskriterier som velegnethet, aksept, naturlighet, behagelighet og tydelighet, bli benyttet (Morton, 1991; Venkatagiri, 1994; Francis, Nusbaum og Fenn, 2007; King, 2007; Campbell, 2007; Taylor, 2009). Blant disse var kriteriet om velegnethet overordnet de andre (Mariniak, 1993; Jekosch, 2005; King, 2007).

I tillegg til ei generell og subjektiv måling, var det ønskelig med ei mer objektiv måling av hvor tydelig Brage kunne sies å være. Blant de forskjellige evalueringsmetodene som fantes var det naturlig å velge mellom metodene som målte tydelighet på setningsnivå. SUS-testen ble valgt til dette formålet (Benoît, Grice og Hazan, 1996). Det var flere fordeler med denne metoden. For det første var ikke testmaterialet ”låst”, dvs. ordnet på forhånd. Dermed kunne jeg utvikle et eget tilpasset testmateriale som gjenspeilet det virkelige produksjonsgrunnlaget til Brage, i stedet for et forhåndsdefinert materiale som kanskje ikke

ville være representativt for Brages bruksområdet i det hele tatt. En annen fordel med SUS-testen var at all semantisk informasjon var fjernet, slik at informantene måtte gjengi det de *faktisk* hørte (Frawley, 1992, s. 12-13). Og selv om testen i utgangspunktet ble designet for å måle talekvalitet på setningsnivå, var det ingenting i veien for å studere datamaterialet både på ord- og fonemnivå i tillegg. Liknende tilnærminger til datamaterialet ble foretatt av Boula de Mareüil m.fl. (2006, s. 1-2), som målte franske talesynteser på ordnivå ved hjelp av SUS-test.

Flere forskere mener at talesynteseteknologien i dag er kommet så langt, at kriteriet om tydelighet synes å bli stadig mindre viktig. Enkelte vil hevde at tydelighetstester kan synes å være unødvendige (Campbell, 2007; Sonntag og Portele, 1998). I denne studien ble tydelighetskriteriet allikevel vurdert som sentralt. Tydelig uttale er en viktig kvalitet for å kunne si noe om hvor velegnet Brage var til å formidle studielitteratur.

Disse to metodene bidro til å belyse forskningsspørsmålene, jf. Kapittel 1.2, s. 2.

4.3 Forskningsetiske vurderinger

Studien er godkjent av Norsk samfunnsvitenskapelig datatjeneste (NSD), og den delen av studien som omhandler brukerundersøkelsen av Datatilsynet. Utvelgelsen av informanter til brukerundersøkelsen ble foretatt ut ifra NLBs lånerregister, som inneholder sensitive personopplysninger som dokumentering av lesehandikap. Disse opplysningene ble behandlet med respekt ved at kun nødvendige opplysninger ble trukket ut. Dette innebærer at studien ikke var anonymisert. Som deltakeridentifisering i datainnsamlinga ble studentenes lånernummer benyttet, slik at ingen andre enn eventuelt ansatte ved NLB skulle kunne være i stand til å identifisere informantene ved navn (Fuglseth og Skogen, 2006).

For det andre ville en kanskje under datainnsamlinga til brukerundersøkelsen komme i dialog med personer som hadde negative opplevelser knyttet til sitt handikap og til temaene som studien berørte (Fuglseth og Skogen, 2006, s. 121). Temaene som ble undersøkt i denne studien var knyttet til enkelte følelsesladde begreper. Det vil for enkelte være knyttet følelser til det å være syns- eller lesehemmet, eller å ha ei funksjonshemming som gjorde at de var avhengig av tilpasset studielitteratur. Dette kan være lett å glemme siden jeg, som ansatt ved Norsk lyd- og blindeskriftbibliotek, arbeider til daglig med tilrettelegging. I verste fall kan

man glemme personene man ”jobber for”. Valget av ei helhetlig evaluering av talekvaliteten til Brage ville naturligvis inkludere den tilrettelagte studielitteratur slik studenten møter den, altså i form av ei DAISY-bok med syntetisk tale (Campbell, 2007, s. 39-41). Tidligere var det vanlig at personer med et dokumentert lesehandikap hadde krav på et hjelpemiddel som kunne spille av DAISY-bøkene. De lesehemmete ble ikke mindre anonyme i klasserommet, i auditoriet eller i lesesalen, med et slikt hjelpemiddel. Nedenfor er det gjengitt ei historie som fanger opp litt av denne problematikken, observert og gjenfortalt av en lærerstudent:

Han hadde fått diagnosen [dysleksi] for flere år siden, og fikk den gang tildelt en bærbar pc. Han nektet å bruke denne fordi han ikke ville bli behandlet noe annerledes enn resten av klassen (Dahlberg, 2009, s. 3).

Selv om historien beskriver situasjonen til en dyslektisk elev i grunnskolen, forteller den allikevel om følelsene enkelte personer kanskje kan ha knyttet til handikapet sitt. Det skal riktignok sies at man som student ved et universitet eller en høyskole skiller seg mindre ut med en bærbar PC i dag, sammenliknet med grunnskolen. Men det kom fram av resultatene til brukerundersøkelsen at ei rekke dyslektiske studenter valgte å benytte seg av iPod eller andre MP3-spillere i stedet for en dedikert DAISY-spiller. Dette er svært uheldig, siden disse studentene mistet muligheten til å ta i bruk alle fordelene som kommer med DAISY-standarden, jf. Kapittel 2.3, s. 5. Dette er et område NLB må undersøke nærmere, for eksempel ved å vurdere muligheten for å tilpasse DAISY-bøkene til lesebrettet, som nå synes stadig å bli mer populært på det kommersielle markedet.

4.4 Brukerundersøkelse

4.4.1 Utforming av spørreskjema

Det ble utformet et spørreskjema, som besto av totalt 66 spørsmål fordelt på seks deler, tilsvarende forslaget fra den utbredte ITU-T-testen, jf. Kapittel 3.5.5, s. 26. Ei slik kategorisk inndeling ville sikre at sentrale evalueringskriterier som tydelighet, naturlighet, behagelighet og velegnethet, ble representert. Første del ble innledet med noen generelle spørsmål for å sikre bakgrunnsvariabler, mens andre del omhandlet stemmekvalitet, tredje del uttale, fjerde del formidling, femte del lesevaner, og sjette del noen generelle spørsmål vedrørende NLBs

produksjon med talesyntese. Femte og sjette del var ikke relevante for denne studien, og vil dermed ikke drøftes her. Undersøkelsen i sin helhet vil bli publisert i egen rapport av NLB.

Spørreskjemaet ble utformet elektronisk og distribuert som hypertekst i henhold til krav om universell utforming (NOU 2005:8, s. 40-41). På denne måten fikk studentene, som ønsket å fylle ut skjemaet på egenhånd, muligheten til det. Men riktignok foretrakk de fleste av studentene i stedet å bli bistått over telefon. Datainnsamlinga over telefon ble gjennomført sammen med en kollega.

4.4.2 Kartlegging av respondenter og datamaterialets størrelse

Tanken bak brukerundersøkelsen var at samtlige aktive studenter med et lesehandikap, som hadde fått studielitteratur som var blitt produsert med Brage, skulle få anledning til å uttale seg om talekvaliteten. Dette inkluderte alle syns- og lesehemmete studenter som tok høyere utdanning, uavhengig om de studerte enkeltemner eller deltok i et studieprogram, og om de fulgte undervisning eller var oppmeldt som privatister. Det var ei utfordring å kartlegge alle aktive studenter siden NLB beregner dette antallet etter hvor mange, som enten i forkant eller i løpet av et semester, har levert inn pensumlistene sine til studenttjenesten eller har vært i dialog med denne. Ei slik beregning ville ikke ta hensyn til de selvstendige lånerne som låner studie bøker uten å være i kontakt med studenttjenesten.

Per 1. juni 2010 var det ved studenttjenesten registrert 184 aktive studenter for vårsemesteret 2010, hvor 43 av dem (23 %) hadde produksjonsrett. Til sammenlikning var det 392 registrerte studenter i løpet av kalenderåret 2009 (Engh og Ellefsen, 2010, s. 18). Det er naturlig å anta at flere av studentene som var aktive i 2009 fremdeles studerte under vårsemesteret 2010. For å belyse problematikken kan det nevnes at bare ved Universitetet i Oslo var det i 2006 om lag 35 synshemmete og 300 dyslektiske studenter, og ved NTNU i 2004 om lag 26 synshemmete og 379 dyslektiske studenter (Kunnskapsdepartementet, 2007, s. 17). Ut ifra en slik antakelse kan det derfor tenkes at det faktiske antallet studenter utgjorde langt mer enn "bare" 184 studenter. Etter å ha studert produksjonsstatistikken fant jeg ytterligere ni lånerne som hadde bestilt studielitteratur med Brage. Ingen av disse var blant de registrerte 184 studentene.

Alle de 184 aktive studentene, i tillegg til de ni som hadde lånt studielitteratur, ble tilsendt invitasjon om å delta både per e-post og som DAISY-bok i vanlig post. Spørreskjemaet ble lagt ved invitasjonen, slik at alle fikk muligheten til å se gjennom spørsmålene i forkant. De som enten ikke var aktive studenter, ikke hadde mottatt studie bøker som var blitt produsert med Brage, eller som ikke ønsket å delta, ble bedt om å gi beskjed om dette. For å unngå et høyt frafall under rekrutteringa til utvalget, ble det vurdert å bruke virkemidler som kunne stimulere til høyere deltakelse, som for eksempel loddtrekning av premier. Men slike virkemidler ble ikke benyttet siden ønsket om godt tilrettelagt studielitteratur måtte antas å være tilstrekkelig for å oppnå høy deltakelse (Fuglseth og Skogen, 2006, s. 134).

Tabell 3. Oversikt over kartlegginga av utvalget til brukerundersøkelsen.

Beskrivelse	Med produksjonsrett	Med lånerett	Ikke registrerte	Totalt
Ikke mottatt studielitteratur med Brage	20	50	4	74
Deltatt i brukerundersøkelsen	18	39	2	59
Ønsket å delta, men leverte ikke innen fristen	2	18		20
Ønsket ikke å delta	1	5		6
Ikke lenger aktiv student		1	1	2
Ikke kommet i kontakt med	2	28	2	32
Totalt	43	141	9	193

Tabell 3 angir at 79 (tjue med produksjonsrett, 57 med lånerett og to uregistrerte) av 85 aktive studenter, som hadde kjennskap til Brage, ønsket å delta i undersøkelsen. Det var ukjent om de 32 studentene som det ikke ble oppnådd kontakt med verken var aktive studenter eller hadde kjennskap til Brage. Etter flere purringer leverte 59 av de 79 studentene, som ønsket å delta, inn skjemaet innenfor fristen. Man må ta høyde for at de 20 studentene som ikke leverte innenfor fristen, kanskje ikke ønsket å delta. Med en slik antakelse ble frafallsraten 30,6 % (59 av 85 studenter deltok). Dette var en akseptabel rate som styrker brukerundersøkelsens reliabilitet. Ifølge Campbell (2007, s. 38) bør et utvalg til slike helhetlige undersøkelser bestå av minst 30 respondenter for at resultatene skal være statistisk signifikante (Skog, 1998, s. 101-106).

Av de 59 skjemaene som ble samlet inn, måtte seks forkastes. Disse var ufullstendige og kunne ikke benyttes. Dermed utgjorde datamaterialet totalt 53 besvarelser, som gir en frafallsrate på 37,6 %. Dette er fremdeles innenfor et akseptabelt nivå. Én respondent leverte inn etter den siste fristen. Denne besvarelsen ble ikke inkludert i datamaterialet.

Data som er framstilt i prosenter, er angitt uten desimaler for å gjøre data mer leselige. Prosentene ble rundet av til nærmeste hele siffer slik at et visst avvik kan forekomme.

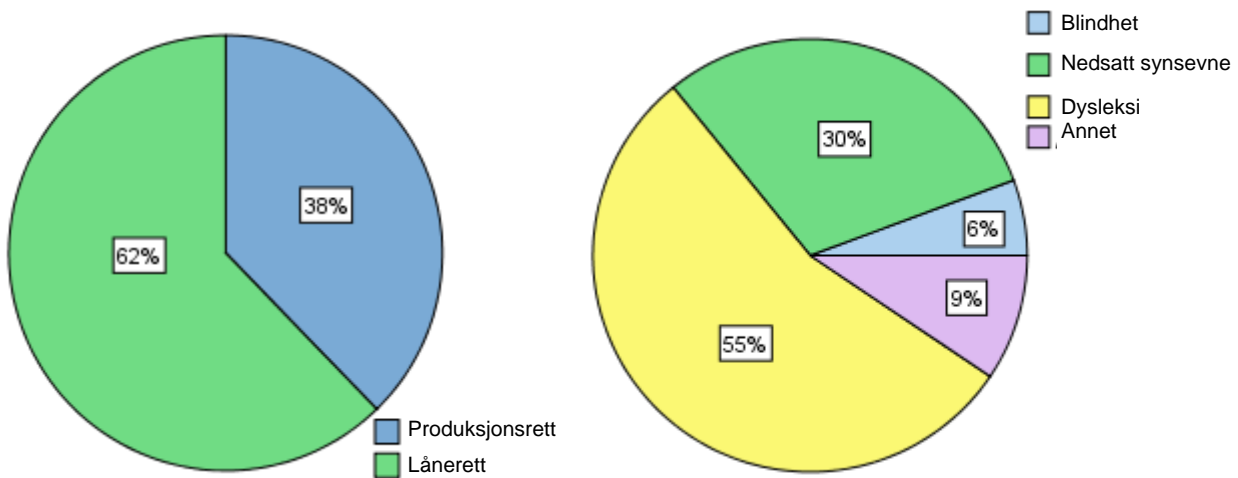
4.4.3 Beskrivelse av respondentene

Det var litt over dobbelt så mange kvinner som menn i utvalget (36 kvinner mot 17 menn), ei fordeling som syntes å samsvare med kjønnsfordelinga blant alle bibliotekets lånere (Sentio, 2008, s. 8). Det var uklart hvorfor de kvinnelige studentene var i klart flertall, men Næss og Støren observerte i 2006 at kvinnene, uavhengig om de hadde et lesehandikap eller ikke, representerte 80 % av veksten i antall nyrekruttede studenter til høyere utdanning i perioden 1991 til 2003 (Næss og Støren, 2006, s. 7). Dersom utvalget fordeles på variabelen alder, var gjennomsnittsalderen blant studentene 33 år. Alderen med flest observasjoner var 22 år. Den yngste studenten var 20 år og den eldste 64 år.

Over halvparten av utvalget besto av dyslektiske studenter. Denne fordelinga etter lesehandikap synes å være unik blant studentene i forhold alle bibliotekets lånere. Ifølge en brukerundersøkelse, som Sentio Research gjorde i samarbeid med NLB i 2008, der formålet var å kartlegge lånerne og se på deres bruksmønster, utgjorde lånergruppa med dyslektikere bare 9 %, mens de synshemmete representerte den desidert største lånergruppa med hele 84 % (Sentio, 2008, s. 10). Selv om enkelte av dyslektikerne hadde utvidet lånerett (2 %), som ga dem samme retten til å bestille obligatorisk pensumlitteratur som studentene med produksjonsrett, måtte de fleste dyslektiske studentene av ressurs hensyn nøye seg med å låne studie bøker som allerede var blitt produsert (Kulturdepartementet, 2005, s. 34).

Det er viktig å være oppmerksom på at dyslektiske studenter vil ha andre behov enn for eksempel blinde eller svaksynte studenter (King, 2007, s. 130-131). Dermed er det nødvendig å fordele utvalget etter lesehandikap.

Figur 2. Fordeling av utvalget etter hvor mange respondenter som hadde produksjonsrett og lånerett (til venstre), og etter lesehandikap (til høyre). Prosent. N=53.



Det kan tenkes at denne situasjonen for dyslektiske studenter påvirket deres holdning til talesyntese generelt, siden denne lånergruppa også må ta til takke med det formatet de får. Dersom en svaksynt student bestiller pensumbøkene sine som talesyntese, kan det hende at dette oppleves frustrerende for dyslektikerne som foretrekker menneskelig tale, og dermed slå uheldig ut på deres bedømmelse av talekvaliteten til Brage. Riktignok oppga de aller fleste dyslektiske studenter at de var svært takknemlig for i det hele tatt å motta deler eller hele pensumet sitt fra NLB. For å illustrere ytterpunktet i denne situasjonen er det nedenfor sitert en av respondentene uten produksjonsrett:

Fordi det ikke fantes bøker i de fagene jeg studerer, ble jeg anbefalt å vurdere et annet studium. (Dyslektisk student om tilgangen på relevant faglitteratur hos NLB)

4.5 SUS-test

4.5.1 Om testen

Selv om et av formålene med SUS-testen var å gjennomføre ei objektiv evaluering av tydeligheten til Brage, hevder enkelte forskere at evaluering av talesyntese aldri kan bli noe annet enn subjektivt (Tomokiyo et al., 2006). Dette må man gjøre seg bevisst før man etterstreber å måle talekvaliteten på en mest mulig objektiv måte. Men selv om det var

utfordrende å måle Filibuster med Brage objektivt, ville en SUS-test sammen med et nøytralt utvalg i alle fall gi mindre subjektive målinger enn ved å velge en annen metode og utvalg.

Dagens talesynteser er av god kvalitet, men er de pålitelige nok til at en lesehemmet student kan stole på at en talesyntese skal være et godt tilretteleggingsalternativ? Som tidligere påpekt hevder enkelte forskere at vi har nådd punktet hvor tydelighet ikke lenger er noe spørsmål, jf. Kapittel 4.2, s. 27. Men resultatene fra brukerundersøkelsen kunne synes å indikere noe annet. Blant de studentene som deltok i brukerundersøkelsen hadde 14 studenter (26 %) liten eller ingen tillit til at Brage leste opp studielitteraturen deres. Så mange som 24 studenter (45 %) foretrakk ikke syntetisk opplesing i det hele tatt.

For å kunne si noe mer om hvor tydelig Brage var, enn det som er mulig å trekke ut ifra funnene fra brukerundersøkelsen, ble det gjennomført en SUS-test med et nøytralt utvalg. I denne SUS-testen ble Brage sammenliknet med to andre talesyntesene. Tanken var å plukke ut de to talesyntesene som respondentene til brukerundersøkelsen oppga som sine favoritter. Men ei uheldig side ved dette var at resultatet blant de norske talesyntesene var svært jevnt fordelt mellom hver stemme. Dermed kunne sammenlikningsgrunnlaget til Brage i SUS-testen synes å være noe tilfeldig. Uansett, to de stemmene Brage ble sammenliknet med var Acapelas Kari og Nuances Stine. NLB benyttet seg av Kari i sin produksjon med talesyntese fra 2004 og fram til Brage ble implementert i produksjonen høsten 2009.

4.5.2 Respondentene

Det kan være store individuelle forskjeller når det gjelder hvilke holdninger lesehemmete studenter har til å motta studielitteratur med talesyntese. Det må tas høyde for at studenter som ikke ønsket å motta studie bøker opplest av Brage, eller som har negative holdninger til bruk av talesyntese generelt, lot dette påvirke på opplevelsen av uttalen til Brage. Med bakgrunn i dette skulle SUS-testen også tjene som ei supplerende måling til undersøkelsen.

Et nøytralt utvalg var ønskelig. Med det menes et utvalg som ikke besto av syns- eller lesehemmete studenter som var avhengig av studielitteratur som produseres med Brage. Utvalget besto av personer som ikke hadde noen bevisst kjennskap til talesyntese eller som ikke var registrert som lånere hos NLB. Utvalget besto av venner og bekjente. Blant de 18

personene som deltok, kom opprinnelig én fra Akershus, fire fra Buskerud, én fra Hordaland (som nå bor i Oslo), én fra Møre og Romsdal, fem fra Oppland, to fra Oslo, tre fra Vestfold (som nå bor i Oslo), og én fra Tyskland (som nå bor i Oslo). Den tyske informanten dro til Norge da hun var 26 år. Hun har nå bodd i Norge i to år og studerer i dag ved en norsk høgskole. Inkludert den tyske informanten var tre av informantene studenter på høgskole- og universitetsnivå, mens resten var i yrkeslivet. Den yngste informanten i utvalget var 27 år og den eldste var 64 år. To av informantene var 27 år, én var 28 år, fire var 29 år, tre var 30 år, to var 32 år, én var 34 år, én var 39 år, én var 54 år, to var 63 år, og én var 64 år. Hensikten med å plukke ut et slikt utvalg var å sikre en mest mulig nøytral tydelighetstest, uten at få et datamateriale som kanskje ville være farget av negative holdninger til talesyntese.

4.5.3 Testmaterialet

SAM-gruppa som utviklet SUS-testen besto av Benoît, Grice og Hazan. I sin rapport fra 1996 listes opp ei rekke krav og restriksjoner for gjennomføringa av metoden og analysering av resultater (Benoît, Grice og Hazan, 1996, s. 381-392). Gruppa utviklet selv programvare for generering av SUS-setninger for nederlandsk, engelsk, fransk, tysk, italiensk og svensk. Jeg kontaktet en av deltakerne i SAM-gruppa, Professor Hazan ved Instituttet for tale, hørsel og fonetikk ved University College London, og fikk et eksemplar av denne programvara av henne. Programvara viste seg å være utviklet på et eldre operativsystem, og kunne derfor ikke kjøres på et nyere. Siden jeg ikke fant noen nyere programvare for å generere SUS-setninger, besluttet jeg å lage programvara selv. Programvara som ble laget besto av tre moduler. Den er tilgjengelig via <http://www.teksttiltale.no/Evalueringsmetoder/SUS/>. Her vil en også kunne finne listene som ble opprettet i forbindelse med SUS-testen.

For å gi en bedre forståelse av hvilke krav og restriksjoner som ble stilt til SUS-testen, vil disse bli gjort rede for i det følgende ved å beskrive modulenes funksjon.

Oppretting av frekvensliste

Det var et krav til SUS-testen at den skulle gjennomføres med de mest frekvente ordene i testspråket. Disse ordene skulle plukkes ut på en slik måte at man sikret de mest høyfrekvente

ordene innenfor hver ordklasse (Benoît, Grice og Hazan, 1996, s. 381-392). Disse ordene skulle bestå av så få stavelser som mulig. For mange språk innebærer dette enstavelsesord. Dersom det ikke forelå tilstrekkelig med enstavelsesord innenfor ei ordklasse, kunne man plukke ut ord med flere stavelser. Alle ordene skulle være meningsbærende. Homonyme forekomster måtte unngås ved at kun den homografen som hadde høyest frekvens, ble brukt.

Den første modulen i programvara imøtekommer disse kravene. Dette ble gjort ved å søke gjennom hele bokbasen til NLB og hente ut tekstinnholdet fra all litteratur på norsk bokmål, produsert med Brage fra og med 1. oktober 2009 til og med 1. september 2010. Hensikten med dette var at jeg ønsket ei frekvensliste som tok utgangspunkt i relevant litteratur i stedet for ei ”vilkårlig” frekvensliste. Tekstmaterialet besto av nesten 24 millioner råord hentet fra ett års produksjon med studie bøker, tidsskrifter og annen litteratur som var blitt produsert med Filibuster med Brage. Som et alternativ til å opprette ei egen frekvensliste kunne en ha valgt ei eksisterende frekvensliste, for eksempel Heggstads *Norsk frekvensordbok* fra 1982. Men ei slik løsning ble valgt bort, ettersom jeg hadde muligheten til å lage min egen ved å ta utgangspunkt i det reelle produksjonsgrunnlaget til Brage. Tanken var at dette skulle gi et datamateriale som gjenspeilte språket som ble behandlet av Brage til daglig.

Oppretting av ordklasselister

Etter at frekvenslista var opprettet var neste trinn å dele de mest høyfrekvente ordene inn i lister fordelt etter hvilken ordklasse hvert ord tilhørte. Dette ble gjort av den andre modulen i programvara. Det ble opprettet totalt 13 lister. Verbene ble delt inn i egne underlister etter modus (imperativ og indikativ), etter om verbene var finitte eller infinitte, og etter transitivitet (transitiv og intransitiv). Hensikten med ei slik underinndeling var å imøtekomme kravet til forskjellige tempora og modi i forhold til de fem syntaktiske strukturene som skulle brukes (Benoît, Grice og Hazan, 1995, s. 384-385).

Videre var det et krav om antall ord per liste. Dette kravet må ses i sammenheng med at det skulle plukkes ut ord til et setningssett med 60 forskjellige setninger. For å unngå at enkelte ord ikke var overrepresenterte i testmaterialet, fastsatte SAM-gruppa ei lavere grense for antall ord per liste.

Fra frekvenslista ble det plukket ut alle enstavelsesord med høy frekvens. Disse ble lagret i ei egen liste og talte totalt 1.140 ord. Som en kuriositet kan det nevnes at disse 1.140 enstavelsesordene utgjorde om lag 0,2 % av den opprinnelige frekvenslista. Dette før alle homografer, duplikatorer, ”skrotord”, enkeltbokstaver, skannefeil og annet ble fjernet. Det var et krav om at kun én homograf skulle beholdes, og det var den med høyest frekvens. For eksempel ble homografer som *lik* og *rask* plukket ut til både substantiv-, verb- og adjektivlista. Siden begge disse homografene hadde høyest frekvens som adjektiv, ble de fjernet fra substantiv- og verblista. ”Skrotord” kan være forårsaket av skrivefeil i den trykte teksten. Og eksempel på skannefeil kan være at adjektivet *lang* og substantivet *jakker* ble elektronisk konvertert til *iang* og *jakkef*. Dette forekom. En siste ting som må nevnes i forbindelse med frekvensutregninga var at frekvensen til hvert ord ble regnet ut på grunnlag av bøyningsforma ordet forelå i i tekstmassen. For eksempel forekom verbet *å få* 31.919 ganger i infinitiv, og 23.274 ganger i presens (får) og 11.426 ganger i preteritum (*fikk*). Og for at ordene skulle plukkes ut til riktig liste, måtte alle ordene tilskrives morfologisk informasjon, blant annet hvilken ordklasse hvert ord tilhørte. Denne informasjonen ble hentet fra Filibusters uttaleleksikon.

Med dette tekstgrunnlaget ble tre av listene utelukkende fylt med enstavelsesord. Dette var listene med adjektiv, substantiv og verb i imperativ. Det var et krav fra SAM-gruppa at kun positive adjektiv skulle plukkes ut til testen, og at disse kun skulle brukes attributivt. Det vil si at de skulle opptre i posisjonen *en ____ kvinne* (Kulbrandstad, 1993, s. 125-129). Dessuten skulle alle nasjonale adjektiv, som *norsk*, *svensk* og *dansk*, fjernes. Det ble også stilt krav til substantivene som skulle plukkes ut. For det første skulle alle låneord fjernes. Dette kravet kunne godt ha vært tydeligere definert, siden det norske språket er ”spekket” med låneord med forskjellig opphav og tid. Kun innarbeidete og alminnelige låneord ble inkludert.

Et moment vedrørende substantivlista må beskrives nærmere. Det har seg nemlig slik at uttaleleksikonet til Filibuster kun skiller mellom utrum og nøytrum, altså ikke mellom de tre genusformene som anvendes i norsk bokmål: *maskulinum*, *femininum* og *nøytrum*. Som sagt ble Filibuster oversatt fra svensk til norsk bokmål. Svensk har som kjent bare to genera: *utrum* og *nøytrum* (Mårtenson og Fjeldstad, 1993, s. 61-63). Denne delen av tilpassinga fra svensk til norsk gjenstår. Det er viktig å understreke at dette ikke har noen praktisk betydning for produksjon av studie bøker og annen litteratur med Filibuster og Brage, siden dette kun berører den morfologiske informasjonen. I uttaleleksikonet er hankjønns- og hunkjønnsord

registrert ortografisk og fonetisk korrekt, dvs. at systemet gjenkjenner maskuline og feminine ordforekomster i en tekst, og tilskriver disse ordene korrekt uttale. Men regelsettet som behandler hankjønns- og hunkjønnsord gjør dette ut ifra ”svensk” morfologisk informasjon. De behandles med andre ord som utrumsord. Det kan også nevnes her at mye av litteratur som produseres er forfattet i et konservativt norsk bokmål, med utrumsformer som *en sol – solen* og *en jente – jenten*. Siden dette ikke har noen praktisk betydning for produksjonen med Brage, er en oversettelse av den morfologiske informasjonen i uttaleleksikonet ei nedprioritert sak. Og det sier seg også selv at med de 731.054 registreringene som i dag utgjør uttaleleksikonet (dvs. hovedleksikonet), vil det være en svært tidkrevende prosess å gå gjennom samtlige utrumsord og skrive dem om til maskulinum og femininum.

Som et resultat av denne ukorrekte morfologiske informasjonen i uttaleleksikonet, kunne det derfor ikke veksles mellom artiklene *en* og *ei*. Med en humoristisk tone kunne en kanskje si at Brage er den første riksmålstaletesyntesen snarere enn en bokmålstaletesyntese.

De resterende listene ble utvidet med både to-, tre- og enkelte firestavelsesord for å imøtekomme kravene om antall ord per liste. Det var for eksempel for få enstavelsesord i verblista til å fylle både lista med transitive og lista med intransitive verb etter SAM-gruppas krav. Dermed måtte det tilføyes en del tostavelsesord, i tillegg til noen trestavelsesord. Siden SUC-tagget som brukes i uttaleleksikonet i dag ikke har definerte tagger for transitivitet, måtte jeg plukke ut disse verbene manuelt. En restriksjon i forhold til de transitive verbene som måtte foretas, var å fjerne verb som tok to objekter siden disse verbene ikke passet til setningsstrukturene, hvor transitive verb var påkrevd (Faarlund, Lie og Vannebo, 1997, s. 663). Andre restriksjoner som måtte foretas i verbene, ikke bare de transitive men alle verbene, var å fjerne hjelpeverb, upersonlige verb og refleksive verb. I spørresetningene var hjelpeverbene *å burde*, *å kunne*, *å måtte*, *å skulle* og *å ville*, nødvendige. Disse ble lagret i ei egen liste slik at dette unntaket ikke berørte noen av de andre syntaktiske strukturene. Det må nevnes at verbene *å klare*, *å komme*, *å lønne* ble beholdt. Disse verbene kan ta et refleksivt pronomen og dermed regnes som refleksive verb. Men siden de også kan opptre syntaktisk uten et refleksivt pronomen, ble de beholdt.

Videre besto konjunksjonlista kun av to konjunksjoner: *eller* og *og*, ettersom det kun var disse to konjunksjonene SAM-gruppa anbefalte å bruke. Neste liste er lista med spørreord. Denne besto også kun av to forekomster: *hvorfor* og *når*. Spørreord tilhører kategorien pro-ord, og regnes ikke som ei egen ordklasse. Pro-ord er ord som ikke har en fast referanse

bestemt ut fra innholdet. Referansen får de i stedet fra konteksten de opptrer i. Pro-ord hører dermed til forskjellige ordklasser. Spørreordene regnes til samme ordklasse som kjernen i den frasetyper de spør etter. For eksempel spør spørreordene *hvem* og *hva* etter en substantivfrase eller et pronomen, og kan dermed regnes som pronomen. I SUS-testen var det et krav om at kun spørreord som regnes som adverb (*hvorfor*, *hvor* og *når*) skulle brukes (Faarlund, Lie og Vannebo, 1997, s. 25-27). Spørreordet *hvor* måtte forkastes siden det ikke passet syntaktisk (*spørreord – verbal – subjekt – direkte objekt*). Den siste lista som måtte opprettes var lista med preposisjoner. Denne var uproblematisk å opprette, og besto av de mest høyfrekvente preposisjonene i norsk bokmål.

Tabell 4 gir en oversikt over listene som ble brukt i SUS-testen. Disse listene er også tilgjengeliggjort gjennom de samme nettsidene hvor programvara er å finne.

Tabell 4. Oversikt over hvor mange oppføringer som kreves innen hver ordklasse og syntaktiske kategori, og hvor mange ord som ble benyttet.

Listekategorier	Navn på liste	Påkrevd antall	Antall benyttet
Substantiv	nouns.lex	> 120	193
Adjektiv	adjectives.lex	> 36	95
Infinitive verb	verbs_infinitive.lex		137
Transitive verb	verbs_transitive.lex	> 48	85
Intransitive verb	verbs_intransitive.lex	> 24	25
Imperative verb	verbs_imperative.lex		70
Hjelpeverb	auxiliaries.lex		5
Preposisjoner	prepositions.lex		24
Spørreord (adverb)	interrogatives.lex	Så mange som mulig	2
Konjunksjoner	conjunctions.lex	Vanligvis 2 (og, eller)	2
Relative pronomen	Tilføyes i SUS-generatoren	Fastsatt i de fleste språk	1

SUS-generatoren

Den siste modulen i programvara var SUS-generatoren. Denne ble brukt til å konstruere setningssettet etter de fem syntaktiske strukturene. De svenske mønstereksemplene i metoden

ble fulgt. Tabell 5 angir de fem syntaktiske strukturene med norske eksempler. A står for Adjektiv, D for Determinativ, H for Hjelpeverb, K for Konjunksjon, P for Preposisjon, RP for Relativ Pronomen, S for Substantiv, Sp for Spørreord (adverb) og V står for Verb.

Tabell 5. De fem syntaktiske strukturene med norske eksempler.

Navn	Struktur	Eksempel (norsk)
1. Intransitiv struktur	D + S + V (intr.) + P + D + A + S	En stol døde til et tomt hus.
2. Transitiv struktur	D + A + S + V (tr.) + D + S	En klar bok sang en ku.
3. Imperativ struktur	V (tr.) + D + S + K + D + S	Se en bok og en sag.
4. Spørrestruktur	Sp + H + D + S + V (tr.) + D + A + S	Hvorfor skulle et tre få et stort brev?
5. Relativ struktur	D + S + V (tr.) + D + S + RP + V (intr.)	En plan spiste en fisk som svømte.

Som eksemplene viser strekker ingen av setningsstrukturene seg over åtte ord. Hensikten med dette var å unngå for lange setninger slik at testen ikke skulle utarte seg som en test i korttidshukommelse. En annen hensikt med dette var muligheten for å få et datamateriale som vil være sammenliknbart på tvers av de fleste europeiske språk (selvfølgelig med enkelte mindre syntaktiske forskjeller) (Benoît, Grice og Hazan, 1995, s. 383-384).

Setningene konstrueres ved at generatoren plukket ut tilfeldige ord fra de listene som hver enkelt struktur påkrevde. Deretter bøyes adjektiv i samsvar med substantiv, og determinativ tilføyes foran substantiv (eller foran adjektiv i tilfeller adjektiv brukes attributivt til et substantiv).

Etter at alle setningene ble ferdig konstruert måtte rekkefølgen av dem stokkes om tilfeldig, slik at informantene ikke skulle kunne gjenkjenne de forskjellige strukturene og bruke dette som hjelpemiddel til å gjette seg fram til utydelige ord. Kort gjengitt var dette SUS-generatorens to eneste oppgaver. Setningssettet som ble generert av denne programvara og benyttet i SUS-testen, er lagt ved bakerst i oppgava.

Det anvendte setningssettet

SAM-gruppa foreslår også hva de opplevde som en anbefalt størrelse på setningssettet som skal brukes i SUS-testen. De foreslo ti setninger innlest med naturlig tale og 60 setninger generert med syntetisk tale, hvor de ti setningene med naturlig tale og ti av de 60 setningene med syntetisk tale skulle fungere som et treningssett. Hensikten med treningssettet var å gjøre informantene bevisst på det språklige og akustiske innholdet som spilles av. Testen ble avgrenset ved å fjerne treningssettet fra testmaterialet. I stedet ble et setningssett med 60 setninger generert ved hjelp av SUS-generatoren. Ei økt varte i om lag 30 minutter, som etter mi mening var nok for å unngå at informantene skulle bli trøtte og få konsentrasjonsvansker.

Setningssettet ble fordelt slik mellom de tre talesyntesene slik at en og en informant hørte 20 setninger med Brage, Kari og Stine. Rekkefølgen på talesyntesene ble variert slik at sammenlikningsgrunnlaget ble likt fordelt blant de 18 informantene.

4.5.4 Gjennomføring

Informantene bosatt i østlandsregionen ble oppsøkt hjemme hos seg selv, mens resten gjennomførte testen da de besøkte Oslo. I forkant av lyttesesjonen ble informantene opplyst om at SUS-testen var en nøytral test hvor de skulle gjengi det de hørte eller trodde de hørte. Ei og ei setning ble spilt av for informantene om gangen. De ble også informert om at setningene ikke var meningsbærende, selv om de besto av virkelige ord. Informantene benyttet hodetelefoner med lydgjengivelse av høy kvalitet (Beyerdynamic DT 770 PRO). Alle talesyntesene ble komprimert og lagret i MP3-format med en bitrate på 160 kbps og en samplingsrate på 22 kHz for ei rettfærdig sammenlikning. Testene ble gjennomført i støyfrie omgivelser med én informant per økt.

4.5.5 Utregning av resultatene

For å sammenlikne skårene mellom de tre talesyntesene, ble hvert ord gitt ett poeng, slik at det var totalt mulig å oppnå sju poeng med de intransitive strukturene, seks poeng med de transitive strukturene, seks poeng med de imperative strukturene, åtte poeng med de

interrogative strukturene, og sju poeng med de relative strukturene. Siden det er tolv setninger av hver struktur var det mulig å skåre maksimalt 408 poeng per informant. Og siden de tolv setningene ble fordelt på seks informanter var det mulig for én talesyntese å skåre maksimalt 2.448 poeng. Ei alternativ poengfordeling kunne vært å gi null poeng dersom ei setning ble gjengitt ukorrekt og ett poeng dersom den ble gjengitt korrekt. Men ei slik poengfordeling sier ingenting om graden av utydelighet eller i verste fall om ei setning er helt uforståelig. Ei poengfordeling på ordnivå vil etter mi mening gi et bedre bilde av balansegangen mellom utydelighet, samtidig som resultatene vil kunne si noe om tydelighet både på ord- og setningsnivå. Poengberegninga var absolutt. Dersom et ord ikke var helt korrekt gjengitt, ble det trukket ett poeng. Dette innebar at dersom informanten gjenga feil artikkel (*et* i stedet for *en*), adjektiv med feil samsvarsbøyning (*frisk* i stedet for *friskt*), eller andre ”småfeil” som å gjengi preposisjonene *mot* og *fra* i stedet for *imot* og *ifra*, ble det trukket poeng. For eksempel i setninga *Når ville et ben leie et greit punkt?* foreslo en informant *Når vil et ben lære et greit punkt?* Man kunne tenke seg å gi ett poeng for ordet *vil* siden det er det samme ordet som *ville*, altså *å ville*, bare i ei anna tid, men siden poengberegninga var absolutt, ble det også trukket et poeng i alle slike tilfeller. Dersom et korrekt gjengitt ord ble slått sammen med et ukorrekt, slik som i setninga *Når ville en lek glede en trygg feil?* der en informant foreslo *Når ville en gløgg glede en trykkfeil?*, ble det gitt verken poeng for *trygg* eller *feil*. Det forekom også tilfeller der informantene ”diktet inn ord”, som for eksempel i setninga *Et bord fulgte et slag som kjente* hvor en informant foreslo *Et bord fullt av et slag som kjente*. Inndiktete ord ble ignorert i poengberegninga, slik at det i dette tilfelle ble det gitt seks poeng av sju mulige (*fulgte* = *fullt av*). I ett tilfelle ble det gjort et unntak; til setninga *Drift en lov og et grep* foreslo en informant *Drift i en lov og et grep*. Dersom *i*-en ble ignorert, ville informanten ha oppnådd seks av seks mulige riktige, men siden setninga måtte oppfattes som ukorrekt gjengitt, ble det allikevel trukket ett poeng.

5 Resultat og diskusjon

5.1.1 Brukerundersøkelse

I det følgende er resultatene fra brukerundersøkelsen framstilt og diskutert. Resultatene er framstilt etter de ulike egenskapene og sidene som ville innvirke på talekvaliteten til Brage.

Syntetisk eller naturlig tale?

Hvor naturlig en talesyntese synes å være må diskuteres i lys av dens prosodiske egenskaper. Tale vil høres mindre naturlig ut dersom den mangler prosodisk variasjon. Dette ga også respondentene uttrykk for i tilbakemeldingene om svakheter ved prosodien til Brage (Morton, 1991, s. 8). Dette gjelder ikke bare for Brage, men for talesynteser generelt. Det er en utbredt enighet blant forskere at to av hovedkriteriene en talesyntese bør måles i forhold til er for det første hvor tydelig den er, og for det andre hvor naturlig den synes å være (Taylor, 2009, s. 523). De fleste er av den oppfatninga at måling av naturlighet foretas i forhold til hvor nært opp til menneskelig tale en talesyntese synes å være (Lampert, 2004, s. 6). Men Campbell (2007, s. 40) foreslår et annet alternativ. I stedet for å måle graden av naturlighet, bør en måle *troverdighet*. Han begrunner dette med å spørre om hvor velegnet ei måling av naturlighet vil være dersom en for eksempel skal måle talekvaliteten til menneskelige ”tegnefilm”-stemmer.

I det følgende vil naturlighetskriteriet bli drøftet i forhold til den utbredte definisjonen. Det vil si hvor nær menneskelig tale respondentene opplevde at talen til Brage var. Deres oppfatning av naturlighet må ses i sammenheng med hvilke erfaringer de hadde med norske talesynteser generelt. Tabell 6 nedenfor viser denne sammenhengen.

Tabell 6. Fordeling av utvalget etter hvor naturlig respondentene syntes Brage var, sammenliknet med en menneskelig innleser, ut ifra hvor lenge respondentene hadde benyttet en norsk talesyntese. N=53.

		ANTALL ÅR						Total
		Mindre enn 1 år	Fra 1-3 år	Fra 4-6 år	Fra 7-9 år	10+ år	Ikke brukt TTS	
GRAD AV NATURLIGHET	Naturlig	2		1				3
	Nokså naturlig	6	2	1	2	4		15
	Nokså unaturlig	8	4	2				14
	Unaturlig	7	1				2	10
	Verken naturlig eller unaturlig	7	2	1		1		11
Total		30	9	5	2	5	2	53

Hele 45 % av respondentene syntes at Brage enten var nokså unaturlig eller unaturlig sammenliknet med menneskelig tale. Det viste seg at ingen av disse 24 respondentene hadde anvendt en norsk talesyntese i mer enn seks år. Til sammenlikning opplevde ingen av dem som hadde brukt en norsk talesyntese i sju år eller lenger at Brage var unaturlig. Denne kommentaren syntes å røre ved kjernen av sammenhengen mellom erfaring og opplevelse av naturlighet:

[...] Toleransegrensa senkes sammenliknet med menneskelig innlesing. Ikke samme forventning til en kunstig stemme. [...] (Blind, 42 år)

Denne respondenten beskrev riktignok sin opplevelse av hvor sentralt behovet for naturlighet var ved å ta utgangspunkt i talekvaliteten til Kari. Men allikevel kan dette resultatet forstås slik at vane syntes å spille ei vesentlig rolle i forhold til hvordan hver enkelt opplevde naturligheten til syntetiske stemmer generelt på samme måte som persepsjonen av syntetisk tale syntes å forbedres over tid (Hjelmquist et al., 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Men det skal også nevnes her at blant dem som opplevde Brage enten som nokså naturlig eller naturlig (34 %), hadde åtte respondenter brukt talesyntese i mindre enn ett år, og fire respondenter fra fire til seks år.

Da respondentene ble spurt om hvor viktig det var for dem at en syntetisk stemme, som skulle benyttes til å produsere studielitteratur, liknet menneskelig tale, oppga 86 % at dette enten var nokså viktig eller viktig. Selv om dette resultatet ikke var overraskende, var

det allikevel viktig å få bekreftet dette. Men for enkelte respondenter var dette verken viktig eller uviktig:

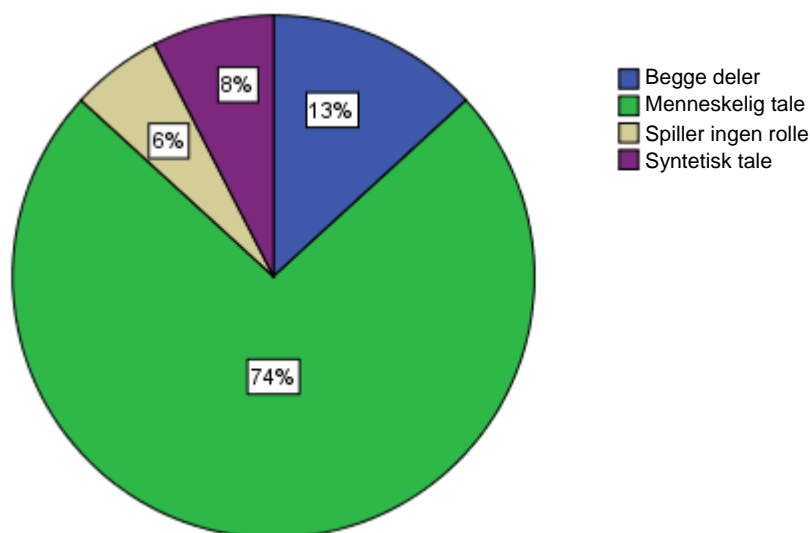
Så er det det ene kontra det andre. Menneskelig tale representerer mer behagelig lyd. Tror jeg synes at det blir viktigere og viktigere med behagelighet. Men jeg har ikke hatt noen problemer med å lese med Kari i timevis. [Jeg har] hatt mye erfaring med å høre på henne i tidsskrifter. [Jeg] kan godt lese tidsskrifter og Aftenposten på PC, men [jeg] bruker DAISY-spiller som avkobling, [siden hun bruker PC mye i jobbsammenheng]. Rytmen i språket er like viktig for behageligheten, men behagelighet er en sammensatt størrelse. (Blind, 42 år)

Denne respondenten brukte syntetisk tale mye i jobbsammenheng, og uttrykte behovet for å få avstand fra syntesen utenom jobbtid. Behovet for å ”koble av” fra syntetisk tale ble også registrert hos Cryer og Home (2009, s. 20-21). Som respondenten oppga, hadde hun ingen problemer med å lytte til en talesyntese i timevis, dvs. Kari, underforstått at bruksverdien kompenserte for en unaturlig talekvalitet. Allikevel konkluderte hun med at graden av behagelighet, som var representert i mindre grad hos en syntetisk stemme sammenliknet med en menneskelig, var å foretrekke. Som hun poengterte, er behagelighet definitivt en sammensatt størrelse, som blant annet er avhengig av prosodiske trekk som leseflyt og rytme. Campbell (2007, s. 41) trekker i tillegg fram lytterens sympati for den syntetiske stemmen som avgjørende for graden av hvor behagelig en talesyntese synes å være for denne lytteren. Med bakgrunn i dette vil graden av behagelighet kunne endre seg over tid, selv for en som beskrev en syntetisk stemme som behagelig første gangen han hørte den.

Men det var også to respondenter som oppga at naturlighetsaspektet i opplesing av studielitteratur var nokså uviktig. Uten å generalisere dette, kan det nevnes at en av dem begrunnet dette med at *til syvende og sist er det ikke stemmen det kommer an på, men at jeg får med meg innholdet*. For denne respondenten var kriteriet om tydelighet viktigere enn kriteriet om naturlighet. Den andre respondenten foretrakk å motta studiebookene sine med syntetisk framfor menneskelig tale. *Greit med syntese for å kunne følge teksten. Veldig greit med tekst når man refererer*. Hun opplevde Brage som unaturlig, men at muligheten til å få tilgang til teksten var viktigere enn behovet for naturlig opplesing. Studiebooker som produseres med Brage synkroniseres med tekstinnholdet i bookene, noe som i dag ikke gjøres med studiebookene som leses inn av profesjonelle innlesere. For denne respondenten var dette åpenbart viktigere enn hvor naturlig Brage var i forhold til en profesjonell innleser. Det kunne også tenkes at ikke alle respondentene kjente til denne muligheten.

Respondentenes opplevelser av hvor naturlig Brage var, kan også ses i sammenheng med om de foretrakk en menneskelig eller en syntetisk stemme til å lese opp studielitteratur:

Figur 3. Fordeling av utvalget etter hva respondentene foretrakk av syntetisk eller menneskelig tale ved opplesing av studielitteratur. Prosent. N=53.



De aller fleste av respondentene (74 %) foretrakk menneskelig tale ved opplesing av studielitteratur. Dette var heller ingen overraskelse siden menneskelig talekvalitet representerer egenskaper som ingen talesyntese i dag er i stand til å gjenskape. De aller dyktigste blant de profesjonelle innleserne vil være i stand til å formidle en fagtekst med engasjement og en naturlig og variert prosodi, og vil dermed kunne holde interessen og ikke minst konsentrasjonen oppe hos lytteren. Dersom dette var standarden respondentene målte naturlighet ut ifra, vil ingen talesyntese være kapabel til å måle seg. Campbell (2007, s. 29-30) advarer også å bedømme naturlighet ut ifra kriterier som vanligvis stilles i bedømmelse av menneskelig tale.

Med bakgrunn i dette forteller disse funnene alene lite om respondentenes opplevelse av talekvaliteten til Brage. Legg for øvrig merke til at det ikke spilte noen rolle for tre av respondentene om de fikk studielitteratur som ble opplest av en syntetisk eller en menneskelig stemme. Det var også interessant at fire av respondentene foretrakk syntetisk framfor menneskelig tale, ettersom denne produksjonsmetoden ga dem elektronisk tilgang til tekstmaterialet. Disse resultatene bør ses i sammenheng med hva respondentene faktisk syntes om det å motta studielitteratur med syntetisk opplesing. I tabell 7 er det forsøkt å generalisere

åpne besvarelser for å gjøre ei statistisk framstilling. Dette kan medføre visse avvik. For eksempel av de 18 som ble plassert i kategorien *Fornøyd med tilbudet*, oppga åtte respondenter at de foretrakk å motta studielitteratur med menneskelig opplesing. Men de syntes likevel å være fornøyd med tilbudet, stort sett ved å begrunne dette med at det var bedre enn ingenting. Tilsvarende funn ble også registrert av Cryer og Home (2009, s. 15).

Jeg synes det er ok med Brage når det ikke går an å få innleste lydbøker. Kari valgte jeg bort, hvis det var mulig. (Svaksynt, 54 år)

Glad så lenge jeg får studielitteratur på lyd. Det letter studiesituasjonen. Det er det viktigste. Man lærer seg å tolke en talesyntese. [...] (Svaksynt, 31 år)

Det er til hjelp. Jeg setter pris på det, selv om det er talesyntese. Men det hjelper på. (Dyslektisk, 23 år)

Jeg synes det er bedre enn ingenting i det hele tatt. Men av og til lurar jeg på om de som har laget talesyntese har satt seg ned og hørt på hvordan det faktisk er å lytte til talesynteseproduksjoner over tid. [...] (Svaksynt, 26 år)

Tabell 7. Fordeling av utvalget etter hva respondentene syntes om å motta studielitteratur som var produsert med syntetisk tale, sammenliknet med hvilket lesehandikap respondentene hadde. Frekvens. N=53.

LESEHANDIKAP					
Opplevelse av tilbudet	Blindhet	Nedsatt synsevne	Dysleksi	Annet	Total
Fornøyd med tilbudet	1	4	10	3	18
Bedre enn ingenting		2	6	1	9
Viktig å få bøkene raskt	1				1
Ønsker tilgang til tekstgrunnlaget	1	1			2
Foretrekker heller menneskelig tale		8	10	1	19
Misfornøyd med tilbudet		1	2		3
Vet ikke			1		1
Total	3	16	29	5	53

Utvalget syntes å være todelt i forhold til hva de syntes om å få studielitteratur med syntetisk opplesing. En unaturlig talekvalitet hos Brage resulterte i at enkelte var misfornøyd og derfor

ikke benyttet seg av tilbudet. Det samme ble bekreftet hos britiske lånere (Cryer og Home, 2009, s. 15).

Liker det ikke, synes det høres uekte ut. (Dyslektisk, 22 år)

Det er ingen erstatning for studielitteratur innlest med menneskelig tale, og for min del klarte jeg ikke å følge med eller lære av lydbøkene der syntetisk tale var brukt. (Dyslektisk, 32 år)

Det har medført at jeg har utsatt studiene. Fikser ikke å lese med talesyntese. Forsøker da heller å lese bøkene [selv], noe som tar tid. (Svaksynt, 48 år)

Det er alvorlig dersom tilbudet med Brage har resultert i at en student måtte utsette studiene sine. Han forklarte videre at han valgte å lese selv siden det å lytte til en talesyntese var helt uaktuelt. Hvorfor er det slik at enkelte opplever det som uproblematisk å lese med talesyntese, mens det for andre resulterer i at studier måtte utsettes?

Det er ikke noe problem. Har man nettopp hørt en menneskelig innlest bok, merker man forskjell. Men det er bare i starten. Etter hvert tenker man ikke på at det er en datastemme. Det går helt greit. (Dyslektisk, 22 år)

For meg personlig er det lettere å lytte avslappet til en vanlig stemme. Jeg greier ikke 100 % å venne meg til en talesyntese. Jeg må lytte mer aktivt for å få med meg innholdet. (Svaksynt, 34 år)

[Det er] tungt når ordene blir feil. Man må lære seg til å høre på det. Man legger merke til det og tenker herregud, går det an?! Men jeg mister ikke sammenhengen. [...]
(Dyslektisk, 35 år)

Dersom trening og vane over tid fører til bedre talepersepsjon av syntetisk tale, bør man forsøke å finne metoder som kan bidra til at flere studenter vil benytte seg av tilbudet (Hjelmquist et al., 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Men dette vil variere fra student til student. Den svaksynte studenten som kommenterte at hun ikke helt klarte å venne seg helt til talesyntese, hadde for eksempel benyttet seg av en norsk talesyntese i mer enn ti år. Hvorfor klarer enkelte å venne seg hurtigere til syntetisk tale enn andre? Kanskje ei løsning kunne ha vært å samle studenter med forskjellig grad av erfaring og opplevelser, og forsøke å skape et forum der studentene kunne ha vekslet erfaringer seg imellom, og ha gitt eller fått råd om effektive studiemetoder med syntetisk opplesing. Hvor mange av studentene i dag kjenner for eksempel til muligheten for elektronisk tilgang til tekstinholdet? For enkelte respondenter gjorde dette studielitteratur med syntetisk opplesing mer verdifull enn med menneskelig opplesing, ettersom dette for eksempel ga muligheten til å

søke etter enkeltord i teksten. Og hvor mange vet at man kan få teksten opplest av en annen talesyntese, dersom man foretrekker andre talesynteser framfor Brage? Som tidligere nevnt syntes det å være flere blant respondentene som ikke hadde tilstrekkelig kjennskap til DAISY-standard, som alle studiebøkene tilrettelegges i, og hvilke muligheter denne standarden gir for syns- og lesehemmete. Blant de 19 respondentene som foretrakk menneskelig tale framfor syntetisk, oppga 13 respondenter at de brukte en DAISY-spiller. Det kan tenkes at bedre kjennskap til DAISY kunne ha endret brukeropplevelsen til de resterende seks respondentene.

Sommeren 2010 ble det besluttet at en forenklet DAISY-spiller, Dolphin EasyReader Express, skulle legges ved alle bøkene som NLB produserer. Alle studenter med en PC med Microsoft Windows operasjonssystem, vil fra dette tidspunktet gis mulighet til å lese studiebøker med en DAISY-spiller. For høstsemesteret 2010 og framtidige studenter kan dette innebære en bedre opplevelse med tilrettelagt studielitteratur. Men det må samtidig understrekes at denne spilleren ikke støttes av Apple Mac- eller Linux-basert operasjonssystem, slik at studentene som bruker dette ikke gis samme mulighet. Et bedre alternativ ville ha vært å legge ved en spiller som støttes av flere operasjonssystem som for eksempel den kryssplattformbaserte spilleren Emerson Reader. Nes (2007, s. 117-118) konkluderte med at spilleren som NLB anbefaler, heller ikke synes å være utviklet med tanke på å imøtekomme behovene til personer med forskjellig lesehandikap.

Men blant de mer eller mindre misfornøyde respondentene, var det også enkelte som ga uttrykk for at de ville forsøke å lese med talesyntese, selv om de syntes det var vanskelig:

Jeg har ikke kastet dem fordi jeg har veldig lyst til å klare å lære meg å bruke dem. [...] Men de hjelper veldig veldig lite, og blir lagt på hylla. [...] (Dyslektisk, 26 år)

Jeg sliter med avkodingen, spesielt innenfor mitt fag [kunsthistorie]. Jeg må avkode mens talesyntesen raser videre. Det blir mer slitsomt. [...] (Svaksynt, 64 år)

Det har medført at jeg har utsatt studiene. Fikser ikke å lese med talesyntese. Forsøker da heller å lese bøkene [selv], noe som tar tid. (Svaksynt, 48 år)

Funnene viste vesentlige forskjeller ved individuelle preferanser, som kommentarene fra de to studentene nedenfor antyder. Siden brukeropplevelsen varierer så mye individuelt sett, vil det være vanskelig å tilfredsstille alle studentene.

Det er helt supert, men det er bedre å øke tempoet med menneskelig tale. Jeg mister lett konsentrasjonen hvis det går for tregt. Men siden jeg ikke har produksjonsrett, er jeg takknemlig for alt jeg kan få. (Dyslektisk, 24 år)

Fint med begge deler [dvs. både menneskelig og syntetisk tale]. Det er lettere å kontrollere tempo med talesyntese. Menneskelig tale har naturligvis en ujevn flyt og tempo, slik at det blir unaturlig og vanskelig å følge i en høyere hastighet. (Dyslektisk, 21 år)

Det var interessant at kun én respondent argumenterte til fordel for syntetisk opplesing. Denne produksjonsmetoden ga henne pensumet hurtigere. Dette var som nevnt ei av hovedårsakene til at NLB besluttet å produsere studielitteratur med talesyntese.

Stemmekvalitet

Under rekrutteringa av innleser til taledatabasen til Brage ble flere krav lagt til grunn. Det skulle være en stemme som "alle" likte (Olaussen og Haugen, 2008). Dette kravet handlet først og fremst om å unngå å velge en innleser med et særegent uttrykk, og i stedet å velge en innleser med jevn, alminnelig stemme. Så mange som 57 % av respondentene likte selve stemmen til Brage godt eller nokså godt. Dette må være en indikator på at selve stemmen syntes å være velegnet til sitt tiltenkte formål, og at talekvaliteten til Brage gir et helhetlig godt førsteinntrykk målt ut ifra det overordnede kriteriet om velegnethet. I tabell 8 blir dette sammenliknet med studentenes erfaring med talesyntese.

Tabell 8. Krysstabell som viser resultatene fordelt på hvor godt respondentene liker selve stemmen til Brage og hvor lenge de har brukt norsk talesyntese. N=53.

ERFARING							
Likte stemmen ...	Mindre enn 1 år	Fra 1-3 år	Fra 4-6 år	Fra 7-9 år	10+ år	Ikke brukt TTS	Total
Godt	5	4	1	2	2		14
Nokså godt	9	2	3		2		16
Verken liker eller misliker	10		1		1	1	13
Mindre godt	3	3					6
Ikke i det hele tatt	3					1	4
Total	30	9	5	2	5	2	53

Det kunne synes å være en viss sammenheng mellom respondentenes kvalitative vurdering av stemmen og erfaring med syntetiske stemmer (Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Blant de som likte stemmen mindre godt eller ikke i det hele tatt, hadde én respondent ikke benyttet seg av talesyntese i det hele tatt, tre respondenter hadde brukt talesyntese mellom ett til tre år, og seks respondenter hadde brukt talesyntese i mindre enn ett år. Av de fire respondentene som ikke likte stemmen til Brage i det hele tatt, hadde én respondent ikke benyttet seg av talesyntese, og tre respondenter brukt talesyntese i mindre enn ett år.

Generelt inntrykk av prosodien

Tabell 9. Resultatene fordelt etter hvilke oppfatninger respondentene hadde om prosodien til Brage (rader), sammenliknet med hvor lenge de hadde benyttet seg av en norsk talesyntese (kolonner). Frekvens. N=53.

		ERFARING						Total
		Mindre enn 1 år	1-3 år	4-6 år	7-9 år	10+ år	Ikke brukt TTS	
PROSODI	Bra	6	1	1		3		11
	Bra når rein tekst			1				1
	Feil ordtrykk og tonelag	4	1	1	1	1		8
	Feil gjengivelse av stavelser	2						2
	Unaturlig rytme	6	4	1	1	1		13
	Generelt dårlig prosodi	1	1				1	3
	Spiller ingen rolle						1	1
	Ingen oppfatning	10	2	1				13
	Vet ikke	1						1
Total		30	9	5	2	5	2	53

Prosodi er viktig for at en talesyntese blant anna skal kunne være i stand til å betone ord, indikere setningsslutt m.m. Dette vil få en syntetisk stemme til å høres mer naturlig ut (Taylor, 2009, s. 13-14). Campbell (2007, s. 32-33) tilføyer karakteristikker som taleridentitet, helse, humør og interesse i forhold til teksten som skal formidles, som også vil berøres av prosodien. Og ettersom prosodien bestemmes ut ifra teksten, vil dette være problematisk for en talesyntese å realisere. Dette resulterer trolig i at slike prosodiske trekk er utelatt i mange

systemer i mindre eller større grad (Taylor, 2009, s. 127-128). Prosodien til en tekstoppassasje som *"Takk for maten!" sa hun og smelte igjen kjøkkendøra*, må defineres av leseren ut ifra den tekstuelle konteksten (Svennevig, 2001, s. 58-63). Selv om slik ekspressivitet vil forekomme oftere i romaner framfor i studie bøker, er det viktig å være oppmerksom på dette.

De fleste av respondentene hadde ingen særlige innvendinger til prosodien til Brage (26 %), og mange oppga at den enten var bra eller akseptabelt (59 %):

[...] Jeg har brukt talesyntese siden 1986. I dag høres det ut som en person. Imponert i forhold til hvordan det en gang var. Det er en helt annen verden. Fra det tidspunktet jeg tok i bruk talesyntese var det bedre enn ingenting. Alle stemmene er i sammenlikning av fantastisk kvalitet. (Blind, 46 år)

Brage er veldig grei i forhold til Nora. Nora var uvant i begynnelsen. Jeg foretrekker østlandsk, altså et mer universelt mål, framfor ålesundsdialekten som skinner igjennom hos Nora. (Dyslektisk, 24 år)

Brage har en bedre setningsmelodi enn Kari. Kari prater mer gebrokkent, dvs. [at hun har] rar setningsmelodi og ordvending. Hun kan for eksempel lese pædagogikk i stedet for pedagogikk. Brage er også bedre enn Kari når det gjelder tonelag, trykk og rytme, også i forhold til Stine, som er veldig monoton og går i den samme melodien hele tiden. (Svaksynt, 47 år)

Siden prosodiske egenskaper som innvirker på talekvaliteten er mange, har jeg i det følgende valgt å diskutere de ulike funnene i hvert sitt underkapittel for oversiktens skyld.

Toneleie og setningsmelodi

Det var enkelte som opplevde at toneleiet til Brage enten var for mørkt (9 %) eller at det var for liten variasjon i setningsmelodien (6 %). Mangel på variasjon i setningsmelodi førte til at enkelte respondenter ble trøtte og utålmodige:

Jeg blir utålmodig og faller lett av. Det er mye pensum jeg skal igjennom. Derfor blir jeg stresset når det går for treigt. (Dyslektisk, 24 år)

Brage leser så sakte at man nesten sovner. [Jeg] gasser opp hvis [teksten] ikke er så interessant eller uviktig. (Blind, 27 år)

Tilsvarende funn ble også registrert hos Cryer og Home (2009, s. 24-25). Årsaka til den monotone setningsmelodien til Brage vil bli nærmere beskrevet i diskusjonen av funnene fra SUS-testen, jf. Kapittel 5.1.2, s. 99-103.

Ordtrykk og tonelag

Prosodiske svakheter relatert til ordtrykk og tonelag ble også trukket fram.

Han gjør for det meste riktig. Noen ganger har han feil trykk, [...], for eksempel ordet grupperinger uttales som gruppe ringer. Men dette er ikke noe problem når jeg har boken foran meg. Jeg humrer litt for meg selv bare. (Dyslektisk, 22 år)

[Prosodien] kunne ha vært gjort bedre. Hadde det vært en person som hadde lest, så ville jo den [personen] ha sett forskjell på tonelag. Det gjør ikke talesyntese. [...] Noen steder klarer den seg helt greit, men andre steder lurar man på hva den driver med. Veldig varierende. (Svaksynt, 26 år)

Brage kan slite med å finne riktig trykk, særlig til fremmedord. (Svaksynt, 45 år)

Svakhetene disse studentene refererte til her, kunne trolig være forårsaket av manglende oppføringer i uttaleleksikonet. Dersom Brage kommer over ordforekomster i en tekst som det ikke er angitt fonetisk transkripsjon til, for eksempel fremmedordene den siste respondenten refererte til, genereres et ordforslag av systemets autogenerator.

Autogeneratoren angir ordtrykk og tonelag etter sannsynlighetsberegninger, som av og til ikke vil være korrekt med uttalenormeringa. Med ordtrykk menes å framheve en stavelse fra de andre stavelsene i samme ord ved å uttale den med større luftpress fra lungene. I ord som har mer enn én stavelse med trykk, fordeles trykket med et hovedtrykk og et bitrykk. Og med tonelag menes tonen eller tonebevegelsen på den trykksterke stavelsen som har til hensikt å gjengi et homonymt ord med en hørbar forskjell fra et anna (Kulbrandstad, 1993, s. 39-41).

Sammensatte ord kan skape vanskeligheter for en talesyntese i forhold til prosodiske egenskaper (Taylor, 2009, s. 116-118). Ta for eksempel det sammensatte substantivet *fagbudsjetter*. Etter norske regler for uttalenormering skal den første stavelsen være trykksterk og uttales med tonelag 2, og den tredje stavelsen ha et bitrykk (Berulfsen, 1969, s. 88).

Dersom Brage skal behandle denne ordforekomsten i en tekst uten fonetisk uttaleinformasjon fra uttaleleksikonet, kan fire mulige uttaleforslag tenkes å bli generert, hvor kun ett er korrekt. For det første kan et rotord som *fag* være utfordrende å angi med korrekt ordtrykk og tonelag,

ettersom det alene skal uttales med tonelag 1, og ved sammenslåing med tonelag 2. Dermed kan den første stavelsen angis med både tonelag 1 og 2. Det norske språket består av ei rekke slike rotord som endrer tonelag ved sammenslåing. For det andre er det i Filibuster tatt høyde for at sammensatte ord utgjør en vesentlig del av ukjente ordforekomster i tekster ved å trene autogeneratoren til å oppdage eventuelle ord slike sammenslåinger kan bestå av. Filibuster kan dermed feilaktig foreslå at det sammensatte ordet *fagbudsjetter* består av det imperative verbet *budsjetter*, ettersom ortografien er identisk med substantivet *budsjetter*. Dermed kan bitrykket i dette eksemplet enten angis ved tredje eller fjerde stavelse.

Det kan også tenkes at feil gjengivelse av ordtrykk kan være forårsaket av at Filibuster plukker ut et feiloppmerket eller -klippet lydsegment når uttale genereres. Dersom lydsegmentet til fonemet /I/ (kort lyd, som i adjektivet *sikker*) er lenket til galt fonem, for eksempel /i:/ (lang lyd, som i substantivet *sik*), kan dette også kunne oppleves som ukorrekt ordtrykk. For eksempel at substantivet *syklistene* uttales med lang /i:/, *sykli:stene*. I slike tilfeller må man se om det er mulig å korrigere lydsegmentene som forårsaker ukorrekt uttale. Dersom det ikke lar seg gjøre, må hele talesekvensen som segmentet ble plukket ut ifra, forkastes.

Slike ordtrykk- og tonelagsrelaterte uttalefeil kan få både smilebånd og nerver til å bevege seg. Selv om det med dagens datamaskiner som blir stadig raskere og får større lagringskapasitet, vil det være en svært omfattende og tidkrevende jobb å føre inn alle nye ordsammenslåinger. Det optimale i slike situasjoner vil være å trene autogeneratoren til å foreslå en uttale med størst mulig treffsikkerhet. Dette innebærer også en tidkrevende jobb med å kartlegge alle tenkelige morfologiske situasjoner som kan forårsake feildefinering av ordtrykk og tonelag. Men i et langtidsperspektiv bør videreutvikling av autogeneratoren i teorien trolig gi det mest optimale resultatet.

Sluking av stavelser

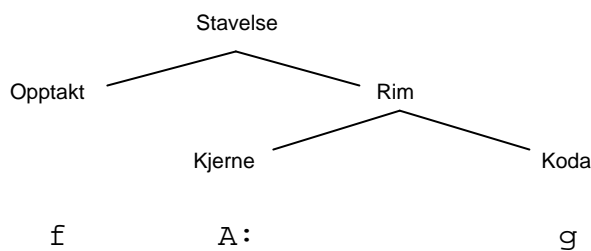
En prosodisk svakhet som enkelte påpekte (4 %), var ukorrekt gjengivelse av stavelser.

Det jeg har hørt av Brage er egentlig ganske bra, men jeg har merket meg at han har en tendens til å sluke eller hoppe over den første delen av enkeltord. Det er vanskelig å dekode ord når man (dette gjelder også folk) unnlater å uttale deler av ord. [...] Dette bidrar

til å oppleve stemmen som mer syntetisk. Men det er altså stort sett i begynnelsen av enkeltord jeg opplever dette hos Brage. (Dyslektisk, 41 år)

For å forstå hva som kan forårsake ukorrekt gjengivelse av stavelser, må det først redegjøres for innholdet i en stavelsesstruktur. Figur 4 angir stavelsesstrukturen til substantivet *fag*.

Figur 4. Eksempel på stavelsesstruktur.



En stavelse kan enten bare bestå av en lyd, en vokal, eller grupperinger med flere lyder rundt vokalen. Dette kalles kjernen. De lydene som opptrer foran kjernen, kalles opptakt, og de bak kjernen, koda. Kjernen og koda utgjør til sammen det som kalles rim (Kristoffersen, Simonsen og Sveen, 2005, s. 188-189). Respondentene som påpekte stavelsesrelaterte uttalefeil, kommenterte at feilen syntes særlig å være gjeldende for opptakta i enkeltord. Selv om enkeltord kan bestå av én eller flere stavelser, ble dette forstått som enstavelsesord. At opptakta i småord kuttet eller slukes, vil lite sannsynlig være forårsaket av mangler i uttaleleksikonet. Det er trolig større sannsynlighet for at slike uttalefeil skyldes feil under den automatiske segmenteringsprosessen av talesekvensene som i dag utgjør stemmen til Brage. Under denne prosessen ble alle setningene eller frasene som ble innlest og lagret i taledatabasen, klippet helt ned til fonemnivå. Alle de 15.604 talesekvensene som i dag utgjør taledatabasen til Brage, er klippet på fonemnivå og utgjør i underkant av 570.000 lydsegmenter. Det er mulig å klippe alle lydene manuelt. Dette vil gi et optimalt resultat, men er til gjengjeld svært tid- og ressurskrevende. Derfor er automatisk segmentering en vel anvendt metode for skjøtesynteser, men en nødvendig bieffekt er visse ukorrekte fonemiske annoteringer (Xu et al., 2005, s. 142). Disse feilene må korrigeres manuelt. Men det er vanskelig å kunne si noe om hvorfor enkelte respondenter opplevde dette særlig i opptakta til

småord. Ei mulig forklaring kan være at enkelte småordene ble lest inn enten for slurvete eller for hurtig, slik at de ikke ble registrert under den automatiske segmenteringa.

Rytme og leseflyt

Flere av respondentene (25 %) opplevde at prosodiske egenskaper som leseflyten og rytmen til Brage syntes å være iørefallende ujevn eller hakkete.

Rytmen liker jeg ikke. Den blir ujevn. Noen ganger glir han og andre ganger ikke. Klart, Brage er bedre enn Kari, men hvis jeg kan bruke Stine, foretrekker jeg heller det. [...] Når setningsrytmen ikke stemmer med det som står skrevet i setningen, da detter jeg ut av og til. (Svaksynt, 50 år)

Litt hakkete, men naturlig for å være talesyntese. [God leseflyt er] Vanskelig å få til med talesyntese. (Dyslektisk, 56 år)

Talesyntese er en vanesak. Man må venne seg til det, så går det bra. (Dyslektisk, 25 år)

Fin prosodi. Bortsett fra noen enkeltord, så er det veldig bra. Helt grei trykk og rytme. Kanskje litt langsomt, og hvis man setter opp hastigheten blir rytmen mer hakkete. (Blind, 27 år)

For den siste av disse respondentene gjaldt det altså særlig dersom lesehastigheten ble økt. Dette er ei utfordring for skjøtesynteser generelt (Xu et al., 2005). Det kunne virke som om de som hadde mer erfaring med talesyntese var blitt vant til dette, siden seks av de tretten respondentene som påpekte dette oppga at de kun hadde brukt en norsk talesyntese i mindre enn ett år, og fire respondenter fra ett til tre år. Som de to dyslektiske respondentene poengterte er god leseflyt vanskelig å få til med en syntetisk stemme, og etter utstrakt bruk av talesyntese over tid kunne man kanskje ha vendt seg til denne sida av talekvaliteten hos en talesyntese (Hjelmquist et al., 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007).

At rytmen til Brage kunne oppleves som ujevn, vil trolig først og fremst være forårsaket av feil under den automatiske segmenteringa. Men at leserytmen syntes å være hakkete, kan være et resultat av valgene som ble tatt under innspillinga av taledatabasen. Det ble besluttet at innleseren skulle lese setningene etter samme toneleiemønster med jevn rytme på hvert enkelt ord, inkludert småord. Det var klart at dette ville resultere i en mindre naturlig opplesing, hvor det ville ha vært større rom for variasjon i setningsmelodien og rytmen

mellom ord som skulle framheves og ikke. Toneleiet og rytmen ble holdt jevnt gjennom hver setning, bortsett fra ved komma hvor innleseren lot toneleiet gå opp, og ved setningsslutt hvor innleseren lot toneleiet gå ned. Da hvert eneste ord ble lest inn med samme jevne rytme, vil dette resultere i en unaturlig og hakkete leserytme. Hensikten med dette mønstret var å sikre størst mulig tydelig gjengivelse av hvert eneste fonem. Dette må ses i sammenheng med at en under utviklinga av Brage satte kriteriet om tydelighet over kriteriet om naturlighet og behagelighet. Konsekvensen av dette valget kan oppsummeres i tilbakemeldingene fra disse tre respondentene:

Det går i én dur. Blir en jevn flyt som man blir sliten av i lengden. (Dyslektisk, 35 år)

I begynnelsen syntes jeg den var håpløs. Flyten er ujevn, som går utover konsentrasjonen. Jeg liker bedre en jevn [talesyntese] uten altfor mye opp og ned. Jeg liker Stine bedre fordi hun er jevnere. (Svaksynt, 50 år)

Noe ujevnt. Av og til trekker den på uttalen og andre ganger lurar man på når han skal bli ferdig med å få uttalt ting. En setning eller et årstall kombinert med dato, er ting Brage sliter med. (Svaksynt, 26 år)

I det følgende vil viktigheten av jevn rytme bli illustrert. *Anerkjennelse av et synspunkt betyr ikke at jeg nødvendigvis er enig i det, eller at det er riktig.* Dette er én av de 15.604 talesekvensene som er lest inn og lagret i taledatabasen til Brage. Ved menneskelig tale kan det forekomme at fonem i småord ikke uttales, som et resultat av vekslende rytme. Dette gjelder særlig dersom hastigheten økes. Det kan i dette eksemplet tenkes at /v/-lyden i det andre ordet, *av*, og /r/-lyden i det tiende ordet, *er*, ved en naturlig rytme uttales svakere eller i verste fall ikke uttales i det hele tatt. I ei innlest studiebok vil dette være tilsynelatende uproblematisk, siden bortfallet av slike lyder gjenspeiles i talemålet. Men i et talesyntesesystem hvor nettopp disse to lydene plukkes ut for å generere uttale til andre ord, kan dette bidra til å gjøre setninga utydelig, i verste fall uforståelig. Prøv for eksempel å tyde innholdet i denne setninga med bortfall av disse to fonemene: *Helse- og sosialarbeidere kan komme i s(v)ært intim kontakt med b(r)ukerne.*

Personer som har brukt talesyntese over tid kan være vant til å dekode slike utydeligheter, men det vil uansett kreve større anstrengelse og konsentrasjon enn dersom samme setning var opplest av et menneske (Axmear et al., 2005).

Som tidligere nevnt har det internasjonale selskapet IBM forsøkt å løse slike ujevnheter med frasespleising (Kvale, 2006, s.22), jf. Kapittel 3.2, s. 10. Det kunne ha vært

spennende å vurdere frasespleising ved å opprette en ekstra taledatabase, der man leste inn de mest frekvente uttrykkene i norsk, som for eksempel fraser som *med hensyn til*, *ved hjelp av*, og *og så videre*. Dersom disse frasene forekommer i en tekst, hentes hele den innleste uttalen til hver frase, i stedet for ordinær enhetsutvelgelse på fonemnivå. Figur 5 illustrerer hvordan uttalen til ei setning genereres av skjøtesynteser.

Figur 5. Eksempel på hvordan Filibuster velger ut lydsegmenter til setninga *Ikke alle tjenester i arbeidslivet er motivert ut fra økonomisk egeninteresse*. Den første linja viser resultatet fra konkateneringa, mens den nederste linja viser hvilke innleste sekvenser hvert segment er hentet fra.

Ikke alle tjenester i arbeidslivet er motivert ut f	ra økonomi	sk eg	eninteresse
<u>Ikke alle tjenester i arbeidslivet er motivert ut fra økonomisk egeninteresse.</u>	Bør vurdere bistand fra økonomisk rådgiver.	Den er ikke normativ, den er en faktisk egenskap.	Ikke alle tjenester i arbeidslivet er motivert ut fra økonomisk egeninteresse.

Ei setning som er lest inn og lagret i sin helhet i taledatabasen, ble prosessert på nytt. Den øverste raden viser at systemet plukket ut sammenhengende fonemer fra tre forskjellige setninger for å generere uttale. Dette blir gjort selv om systemet kunne ha valgt å hente fram setninga i sin helhet. Som et tiltak for å jevne ut rytmen og setningsmelodien til Brage, kan frasespleising vurderes dersom dette viser seg å gi gode resultater (Kvale, 2006, s.22).

Enkelte av respondentene bemerket at rytmen ble mer hakkete desto lengre setninger Brage skulle håndtere, og andre opplevde at Brage klarte seg bra i noen bøker, mens det hakkete mer i andre:

Litt hakkete når setningene blir lange. [Brage] Blir litt forvirra, særlig ved komma. (Svaksynt, 30 år)

På de beste bøkene er flyten ganske bra, men på andre bøker hakker den mer. Det varierer. (Svaksynt, 64 år)

At Brage syntes å tolke og behandle visse bøker bedre enn andre kan trolig skyldes at uttaleleksikonet til Filibuster inneholdt flere av fagtermene som var representative for ei bok sammenliknet med ei anna. Å sikre et akseptabelt nivå av tydelig gjengivelse av sentral fagterminologi innenfor hver enkelt studieretning, er naturligvis et av de viktigste områdene

som må prioriteres. Dette kan bli ei utfordring i forhold til tids- og ressursaspektet, ettersom det vil være naturlig å prioritere de studieretningene med flest aktive studenter.

Lesehastighet

Måling av lesehastighet er et sentralt moment i forhold til å bruke talesyntese som hjelpemiddel i studiesammenheng. Det er gjennomført studier som indikerer at høy lesehastighet vil kunne bidra til å gjøre en syntetisk stemme mindre tydelig og i verste fall uforståelig (Koul, 2003; Kalvik et al., 2010). Å måle lesehastighet som en innvirkende faktor på talekvaliteten er ikke uproblematisk, ettersom det synes å være stor variasjon med individuelle preferanser i forhold til lesehastighet. Hjelmquist, Dahlstrand og Hedelin (sitert i Cryer og Home, 2008, s. 8-9) gjennomførte i 1992 en studie, hvor det ble antydnet å være en sammenheng mellom lesehastighet og erfaring med bruk av talesyntese. Ifølge denne studien syntes blinde og svaksynte respondenter med mindre erfaring å foretrekke en lavere lesehastighet, og tilsvarende ønsker de med mer erfaring å øke lesehastigheten.

Respondentene som benyttet seg av en DAISY-spiller, vil kunne ha muligheten til å justere lesehastigheten opp og ned etter eget ønske (Kerscher, Kahl og Sutton, 2008). Det er aspekter ved slik opp- eller nedjustering som bør diskuteres. Men først må det gjøres rede for lesehastigheten til Brage, som synes å stå i en særstilling til øvrige stemmer. I Tabell 10 nedenfor er Brage sammenliknet med andre norske stemmer. Tabellen viser tida hver enkelt talesyntese brukte for å lese en moderat lang tekst. Tekstgrunnlaget som ble benyttet i sammenlikninga var en wikipediaartikkel om Ivar Aasen, bestående av 2.585 ord (13.403 tegn uten mellomrom). Lenke til teksten kan finnes i litteraturlista. Dette var en bokmålstekst med flere tekstuelle utfordringer som innslag med riksmål, årstall, forkortelser m.m. Tilpasninger ble gjort for å gjøre sammenlikningsgrunnlaget mest mulig likt. For eksempel er parenteser i teksten erstattet med ordet ”parentes” foran en parentes, og ”parentes slutt” etter parentesens slutt. Hensikten med denne utskiftninga var at en serverbasert talesyntese, som NLB Brage, vil erstatte parenteser og sitater automatisk, noe som ikke gjøres av klientbaserte stemmer. Målingene ble gjort ut ifra talesyntesenes normalhastighet.

Tabell 10. Oversikt over differansen i spilletid mellom Brage og andre norske kommersielle talesynteser. Målingene er gjort ut ifra talesyntesenes normalhastighet.

Talesyntese	Total spilletid	Differanse i minutter	Differanse i prosent
NLB Brage	26:35		
Acapela Kari	18:26	8:09	30,66 %
Acapela Olav	18:27	8:08	30,60 %
Acapela Trygve	19:29	7:06	26,71 %
Acapela Vegard	19:29	7:06	26,71 %
Nuance Nora	20:23	6:12	23,32 %
Nuance Stine	17:06	9:29	35,67 %
Isak HTS	18:33	8:02	30,22 %
LingSpeak Arne	15:39	10:56	41,13 %

Som tabell 10 viser, var den gjennomsnittlige spilletida for disse ni talesyntesene, nitten minutter og tjueen sekunder. Brage lå hele sju minutter og fjorten sekunder (27,21 %) over gjennomsnittet, som utvilsomt innebærer at Brage er den tregeste blant de norske talesyntesene. I ei slik sammenlikning er det viktig å bemerke at differansen angir den totale spilletida, noe som ikke nødvendigvis utelukkende vil være identisk med den *faktiske* lesehastigheten. Faktorer som ikke kan leses ut ifra disse tallene, er hvert enkelt systems forskjellig grad (og løsning) av tekstforprosessering, som vil bidra til å forlenge eller redusere den totale spilletida. For eksempel uttalte Brage forkortelsen *utg.* som *utgitt*, mens Kari uttalte dette som et ukjent enstavelsesord. Et annet eksempel er at Isak HTS ikke leste ekspansjonen *til* i årsintervallet 1957-60. Brage leste dette som *nittenhundreogfemtisju til seksti*. Alle slike forskjellige tekstforprosesseringsløsninger vil bidra til å påvirke den totale spilletida.

Sammenlikninga kunne også ha blitt gjort med en lettfattelig tekst som ikke krevde noe særlig tekstnormalisering, slik at lesehastigheten ble sammenliknet uten å trekke inn slike tekstuelle utfordringer. Dette var ikke ønskelig, ettersom en lettfattelig tekst ikke ville ha representert den brukeropplevde lesehastigheten slik studenter møter den i studielitteratur.

Med bakgrunn i denne sammenlikninga av lesehastighet, vil respondentenes opplevelse av lesehastigheten til Brage i normalhastighet diskuteres i det følgende. I tabell 11

er lesehastigheten sammenliknet med hvor lett respondentene syntes det var å få med seg meningsinnholdet i tekster, som ble opplest av Brage.

Tabell 11. Fordeling av utvalget etter hva respondentene syntes om lesehastigheten til Brage i normalhastighet (kolonner), sammenliknet med hvordan de syntes det var å få med seg meningsinnholdet i en tekst, som ble opplest av Brage (rader). N=53.

LESEHASTIGHET							
FORSTÅELSE		Langsomt	Nokså langsomt	Passelig	Nokså raskt	Raskt	Total
	Lett	3	2	5			10
	Nokså lett	2	5	7	1		15
	Nokså vanskelig		4	5		1	10
	Vanskelig	1	1	3	1		6
	Verken lett eller vanskelig	3	2	6	1		
Total	9	14	26	3	1	53	

Studentene syntes å være todelt i forhold til oppfatninga av lesehastigheten til Brage. 49 % av respondentene syntes at den var passelig, mens 43 % syntes den enten var langsom eller nokså langsom. Blant de seks respondentene som oppga at de syntes det var vanskelig å få med seg meningsinnholdet, var det tre respondenter som opplevde at lesehastigheten var passelig, to som syntes den enten var langsom eller nokså langsom, og én som syntes den var nokså rask. Kun én respondent syntes at lesehastigheten til Brage var rask. Hun oppga også at hun syntes det var nokså vanskelig å få med seg meningsinnholdet i en tekst som Brage leste.

Besvarelsene nedenfor illustrerer den varierende opplevelsen av lesehastigheten:

Jeg velger å lese på egenhånd, fordi det går så fort med Brage, [...] (Svaksynt, 22 år)

Jeg må la talesyntesen snakke sakte for å få med meg alt, og for å kunne følge med i teksten og skrive notater. Stemmen ble litt vrent og rar [i høyere hastighet]. [...]. I normalhastighet snakker han [Brage] veldig fort. (Dyslektisk, 30 år)

[...] Jeg går ned i tempo, men aldri med Brage. Hvis det er vanskelig stoff eller hvis jeg tar notater, går jeg helt ned på det laveste hakket. Men Brage endrer jeg aldri på. Jeg går så godt som aldri opp i tempo. (Svaksynt, 50 år)

Han [Brage] er såpass sakte og tydelig, slik at man kan tenke seg til ord som er utydelig uttalt. Men med Kari rekker man ikke å tenke seg til det. (Dyslektisk, 22 år)

[...] Brage er virkelig bra. Jeg kjørte ham skikkelig og ble virkelig positivt overrasket. Min tidligere erfaring med andre talesynteser som ikke har klart dette [dvs. å gjengi teksten tydelig i svært høy hastighet] like bra, gjorde at jeg ikke har brukt denne funksjonen [dvs. justering av hastigheten] før med Brage. (Svaksynt, 34 år)

[...] Men da jeg fikk Brage tenkte jeg Hjelp, nå må jeg laste Vegard inn igjen. Jeg blir irritert av utgangspunktet for hastigheten, fordi det går altfor sakte. Og jeg blir sliten når jeg øker tempo, fordi det blir for stakkato. Jeg er vant til å lese fort. Jeg har ikke tid til å lese så sakte. Det blir for teit at det skal gå så langsomt. Må rasjonaliseres! (Blind, 42 år)

Brage leser så sakte at man nesten sovner. (Blind, 27 år)

Det er oppsiktsvekkende at opplevelsen av lesehastigheten var så forskjellig hos respondentene. Jeg hadde antatt at langt flere skulle oppleve hastigheten som langsom, ettersom det på forhånd var kjent hvor sakte Brage leser, sammenliknet med alle andre norske talesynteser. NLB vurderte i forkant av undersøkelsen å justere normalutgangspunktet for lesehastigheten til Brage noe opp, men ut ifra disse funnene bør hastigheten forbli uendret. Siden lesehastigheten til Brage skiller seg så betraktelig fra hastigheten til de øvrige norske talesyntesene, bør denne unike egenskapen til Brage beholdes. På denne måten kan de som har behov for en tregere talesyntese få det, mens studenter som heller foretrekker en hurtigere talesyntese, kan selv velge dette etter eget ønske. Men hvilke talesynteser foretrakk respondenter som syntes Brage leste sakte, og hvilke foretrakk de som syntes Brage leste raskt? Tabell 12 nedenfor viser ei slik fordeling av resultatene.

Tabell 12. Fordeling av utvalget etter hva respondentene syntes om lesehastigheten til Brage i normalhastighet (rader), sammenliknet med hvilken norsk talesyntese de foretrakk (kolonner). N=53.

NORSKE TALESYNTESER											
LESEHASTIGHET		Brage	eSpeak	Kari	Nora	Stine	Vegard	andre	Ikke TTS	Ubesvart	Total
	Langsomt	2			2		2		3		9
	Nokså langsomt	5		2				1	6		14
	Passelig	7	1	1		2		1	14		26
	Nokså raskt					1			1	1	3
	Raskt	1									1
Total	15	1	3	2	3	2	2	24	1	53	

Denne sammenlikninga av resultatene viser at blant de 23 respondentene som syntes lesehastigheten til Brage enten var langsam eller nokså langsam, foretrakk sju av disse (30 %) Brage som oppleser av studielitteratur. Det gjorde også den éne respondenten som syntes lesehastigheten var rask. Men det er ukjent om denne respondenten foretrakk Brage siden de andre norske talesyntesene leser enda raskere.

Dette må forstås slik at ingen behov er like. Noen behøver en talesyntese man kan lese sakte med, ta notater med, er så saktelesende at det er mulig å tyde selv utydelige ord, mens andre behøver en talesyntese som de kan lese raskt med, ikke miste konsentrasjon eller bli irritert av – og ikke sovne av.

Bruk av pauser

Innslag med pauser vil kunne være av betydning for hvor naturlig respondentene opplevde sider ved talekvaliteten som leseflyt og rytme. All litteratur, enten om den produseres med syntetisk eller menneskelig tale, vil inneholde pauser, som i denne sammenhengen vil si korte periodiske innslag med stillhet. I tilrettelagt litteratur brukes pauser for å markere overganger mellom de forskjellige tekstuelle nivåene, som for eksempel kapitler, avsnitt, utdrag, sitater, parenteser, etc. Mellom de øverste tekstnivåene vil det være lengre pause enn ved de laveste tekstnivåene.

72 % av respondentene syntes studielitteraturen som ble lest opp av Brage inneholdt en passelig mengde med pauser, mens 17 % syntes det enten var for få eller nokså få, og 12 % enten for mange eller nokså mange. Tilsvarende opplevde 75 % av respondentene at lengda på hver enkelt pause var passelig, mens 11 % syntes pausene enten var for korte eller nokså korte, og 13 % enten for lange eller nokså lange. Det syntes ikke å være noen sammenheng mellom opplevelsen av verken mengden med pauseinnslag eller lengda av pausene, og lesehastigheten. Da en vesentlig stor andel av utvalget ikke hadde noe å utsette på bruken av pauser, må disse variablene forstås som generelt akseptable.

Uttale

Måling av uttalen til Brage måles ut ifra kriteriet om tydelighet. Den vanligste måten å måle hvor tydelig en syntetisk stemme synes å være, er med tester som tar i bruk ”tulleord” eller setninger uten noe meningsinnholdet, slik som SUS-testen som ble anvendt i denne studien. Campbell (2007, s. 40) mener forskere er for opptatte med å måle ytelse framfor anvendbar pålitelighet og funksjonsstabilitet, som slike metoder ikke tar høyde for. Den ideelle målinga av tydelighet ville for Campbell ha vært i en situasjon med faktisk bruk, for eksempel med et testmateriale bestående av pensumbøkene til hver enkelt student og gjerne i det miljøet hvor studentene studerer. Som en forstår vil dette være nærmest umulig å administrere. De ulike miljøene kan være studenthybler, lesesaler, men også i bilen på vei til lærestedet, eller på bussen, ettersom det er kjent at enkelte av respondentene foretrakk å kopiere over bøkene til sin iPod eller MP3-spiller, og lytte mens de er i bevegelse. Og testmaterialet vil også være utfordrende å vurdere siden innhold og mengde vil være forskjellig fra pensum til pensum. Riktignok er forslaget til Campbell en tankevekker, men vil være altfor utfordrende, kostbar og tidkrevende både å gjennomføre og administrere.

Selv om det kan være delte meninger om hvilken metode som gir det beste resultatet for ei måling etter kriteriet om tydelighet, er det i alle fall en utbredt enighet blant forskerne om at dette evalueringskriteriet er et av de to viktigste, ved sida av naturlighet (Taylor, 2009, s. 523). I det følgende vil respondentenes opplevelse av hvor tydelig Brage var, bli diskutert ut ifra ulike sider ved uttalen til denne talesyntesen.

Tabell 13. Fordeling av utvalget etter hvor tydelig studentene syntes den generelle artikuleringa til Brage var, sammenliknet med hvilken norsk talesyntese de foretrakk som oppleser av studielitteratur. N=53.

TALESYNTSE										
ARTIKULERING	Brage	eSpeak	Kari	Nora	Stine	Vegard	andre	Foretrekker ikke	Ubesvart	Total
	Tydelig	10		1		1		8		20
	Nokså tydelig	5	3		3	1	1	10		23
	Nokså utydelig		1	1			1	6	1	10
Total	15	1	3	2	3	2	2	24	1	53

Hele 43 av 53 respondenter (81 %) syntes at Brage enten hadde ei tydelig eller nokså tydelig artikulering. Ingen oppga at de opplevde den som utydelig. Av de resterende ti studentene som syntes den generelle artikuleringa var nokså utydelig, hadde fem benyttet seg av en norsk talesyntese i mindre enn ett år og to hadde ikke benyttet en norsk talesyntese i det hele tatt. Disse funnene bekrefter igjen at vane og tilvending av syntetisk tale bidrar til å forbedre forståelsen over tid (Hjelmquist et al., 1990; Vankatagiri, 1994). En kan også forstå dette funnet ut ifra studentenes individuelle preferanser, siden samtlige som opplevde artikuleringa til Brage som nokså tydelig, heller foretrakk andre talesynteser (eller menneskelig tale) framfor Brage. Respondentene med større erfaring foretrakk talesyntesene de selv var vant til:

[...] Det er Vegard jeg er vant til. Jeg har jobbet lite med å vurdere talesynteser seg imellom. [...] Erfaringen med å høre tidsskrifter med Kari-stemmen, er at jeg har blitt mindre sliten sammenliknet med å høre med Vegard. [Men allikevel foretrakk respondenten Vegard]. Man er litt vanedyr! (Blind, 42 år)

Det er også interessant å understreke at så mange som 18 stykker (75 %) av de respondentene som ikke foretrakk studielitteratur med syntetisk opplesing, allikevel syntes at den generelle artikuleringa til Brage var enten tydelig eller nokså tydelig. Alt i alt syntes dette å gi et generelt godt helhetlig inntrykk av hvor tydelig respondentene syntes Brage å være.

Tabell 14. Fordeling av utvalget etter hvor ofte respondentene syntes uttalemessige utydeligheter forekom. Frekvens. N=53.

		HYPPIGHET				
		Ofte	Av og til	Sjelden	Aldri	Ubesvart
UTTALEMESSIGE UTYDELIGHETER	Feil uttale av et ord slik at ordet blir uforståelig?	8	23	19	3	
	Feil uttale av et ord slik at setninga blir uforståelig?	6	13	24	10	
	Kutting eller sluking av ord?	4	28	15	6	
	Feil uttale av sammensatte ord?	9	24	15	5	
	Feil betoning av homografer?	5	23	18	6	1
	Uforståelig eller feil uttale av tall og talluttrykk?	12	11	22	8	
	Uforståelig eller feil uttale av norske egennavn?	4	20	21	8	
	Uforståelig eller feil uttale av utenlandske egennavn?	13	18	19	3	
Totalt (prosentvis)		14,4 %	37,7 %	36,1 %	11,6 %	0,2 %

En feil kan være forårsaket av forskjellige svakheter i de forskjellige delene av systemet. I det følgende er det allikevel forsøkt å peke på feilkilden til hvert enkelt tilfelle. Men først må det poengteres at generelt sett var det ifølge respondentene særlig sammensatte ord, tall og talluttrykk, og utenlandske egennavn som ofte syntes å være utydelig uttalt. For det andre kan en også legge merke til at respondentene opplevde at Brage oftere var mer utydelig på ordnivå enn på setningsnivå. Dette kan forstås slik at respondentene kunne danne seg mening i utydelige setninger, kanskje ut ifra konteksten, selv om ei setning inneholdt et utydelig ord.

Utydelig uttale av enkeltord kan for eksempel være forårsaket av at det gjeldende ordet ikke er registrert i uttaleleksikonet, og at uttaleforslaget som autogeneratoren foreslår, vil være ukorrekt. Uttaleforslaget vil være bygget på norske ortografiske regler. Med bakgrunn i dette vil det være forståelig at respondentenes opplevelse av uttale av utenlandske egennavn, som ikke var oppført i uttaleleksikonet, ofte kunne være vanskelig å tyde. Enkelte egennavn vil være utfordrende å uttale korrekt, selv for en menneskelig innleser. Som nevnt kan man ikke gjette seg fram til riktig uttale av egennavn som *Leicester*, *Edinburgh* og *Begin*, bare ved å se på ortografien. Man må rett og slett vite hvordan egennavnene skal uttales.

Men utydelig uttale av enkeltord kan også være forårsaket av feilklipping under den automatiske segmenteringsprosessen. Dette må antas å være gjeldende særlig hvis respondentene opplevde at ord ble ”kuttet” eller at stavelser ble ”slukt”. I slike tilfeller kan dette skyldes at en eller flere lyder i et ord er konkatenerert med lydsegmenter, som enten ble svakt innlest under innspillinga av taledatabasen til Brage, eller som er blitt feiloppmerket, jf. Kapittel 5.1.1, s. 56-57.

Mulige årsaker til ukorrekt uttale er tidligere også drøftet i forhold til ordtrykk og tonelag, både ved tilfeller med sammensatte ord og med homografer, jf. Kapittel 5.1.1, s. 55. Dette er ei særlig utfordring for en talesyntese. Ta for eksempel mange norske preteritumsverb som representerer ei homonym utfordring: *arbeidet*, *badet*, *funnet*, *landet*, *siktet*. Felles for disse er at de ender på *-et* og kan enten være et substantiv eller et verb. Filibuster må i slike tilfeller definere korrekt ordtrykk og tonelag ved å studere konteksten ordene opptrer i ut ifra den ortografiske og morfologiske informasjonen disse kontekstuelle ordene gir. En slik regel kunne for eksempel ha vært utformet slik at dersom den ortografiske forekomsten *budsjetter* opptrer i en tekst, og den neste ordforekomsten er morfologisk angitt som et verb, ville Filibuster tolke forekomsten *budsjetter* som et substantiv, som for eksempel i denne setninga: *Statens budsjetter kan diskuteres*. På tilsvarende måte ville Filibuster ha

tolket den ortografiske forekomsten *landet* som et verb dersom ortografien til den forrige ordforekomsten enten er *ha, har, hadde, bli, blir, ble, blitt*. Men dette er ingen optimal løsning. Dersom det skal lages slike regler med ortografisk framfor morfologisk informasjon, innebærer dette at det må lages en regel for hver eneste ortografiske ordforekomst. Bare homonyme preteritumsverb i norsk bokmål alene vil utgjøre ei lang liste. Problemet i dette tilfellet er at det i dag ikke er opprettet noen morfologisk tagg for å sonde hjelpeverb fra øvrige verb. Kunne man ha skrevet én generell regel som kunne ha fanget opp alle slike tilfeller, i stedet for å lage en ny regel for hver eneste ortografiske ordforekomst, ville man for eksempel ha sluppet ei lang og lite lesbar rekke med ortografiske unntaksregler for norske homonyme preteritumsverb. Det er ei rekke flere regler som kunne ha vært skrevet generelt ut ifra hjelpeverbenes kontekstkrav (Åfarli og Eide, 2003, s. 131-142).

Når det gjelder tall og talluttrykk vil ukorrekt uttale av dette vanligvis skyldes svakheter eller mangler i tekstforprosessorens regelsett. Det er for eksempel laget regler for å kunne sonde mellom en dato, et valutabeløp, et telefonnummer, et årstall, og ei rekke ulike numeriske intervaller. Slike regler er utformet ved å la Filibuster studere konteksten slike tall og talluttrykk opptrer i, og ikke minst hvordan tallet og talluttrykket er skrevet.

Alle disse regler og registreringer av nye ord og egennavn må foretas etter norske uttaleregler, som i seg selv er ei meget utfordrende oppgave. Skal Brage for eksempel uttale året *2010* som *totusenogti, tjueti, tjuehundreogti* eller *tyvehundreogti*? Ettersom det i dag ikke er utarbeidet noen retningslinjer for hvilke språklige føringer som skal være gjeldende for Brage, følges uttaleforslagene til Berulfsen (1969). Men dette er helt klart en diskusjon som bør tas i et forum hvor både Nasjonalbiblioteket, som i dag administrerer og videreutvikler Norsk språkbank, dvs. uttaleleksikonet som benyttes i Filibuster, Norsk språkråd og andre sentrale språkaktører, bør være representerte.

Tolking og behandling av fagterminologi

Studieretningene representert med størst andel av respondenter var utdanningsvitenskapelige (26 %) og samfunnsvitenskapelige studier (21 %), etterfulgt av humanistiske (13 %) og medisinske fag (13 %). Som tabell 15 nedenfor angir, oppga 40 % av respondentene at de syntes Brage leste faguttrykk innenfor deres studieretning på en ok måte. Respondenten som

ikke oppga noe svar, begrunnet dette med at hun ikke hadde tilstrekkelig erfaring med Brage, siden hun valgte å bruke en annen talesyntese til å lese opp studielitteraturen.

Tabell 15. Oversikt over hvordan respondentene opplevde at Brage mestret faguttrykk innenfor deres studieretning, sammenliknet med hvilken studieretning respondentene deltok innenfor.

MESTRING AV FAGTERMINOLOGI							
Studieretning	Godt	Nokså godt	Ok	Nokså dårlig	Dårlig	Ikke oppgitt	Total
Ex. Phil./Ex. Phac.		1					1
Humanistiske fag	1		4	1	1		7
Jus				1	1		2
Matematikk og naturvitenskap			3	1			4
Medisin		2	3	2			7
Samfunnsvitenskap		3	2	3	3		11
Tverrfaglig			1				1
Utdanningsvitenskap	2	3	7	2			14
Videreutdanning			1				1
Økonomi, administrasjon og ledelse	1			1	2		4
Annet						1	1
Total	4	9	21	11	7	1	53

Det må her tas høyde for at NLB kan ha valgt å produsere enkelte studiebøker som var avhengig av profesjonell menneskelig opplesing for å kunne formidles på en akseptabel måte:

Noe av studielitteraturen jeg fikk denne gangen [dvs. i løpet av vårsemesteret 2010], altså tegnspråk, går rett og slett ikke med talesyntese. Rein historikk går greit, men det nytter ikke å forklare den praktiske biten ved tegnspråk med en talesyntese. (Svaksynt, 34 år)

Tilfellet som respondenten her kommenterte er beklagelig. Studiebøker om tegnspråk hvor figurer vil utgjøre en naturlig stor andel av bokinnholdet, kan ikke produseres med talesyntese. Disse bøkene må formidles av en fagperson som kan tegnspråk og vet hvordan dette best mulig kan formidles.

Videre var det interessant at én av de fire respondentene som syntes Brage mestret faguttrykk på en god måte, studerte økonomi, administrasjon og ledelse. Denne studieretninga omfatter pensumbøker som vil kreve mye tekstforprosessering, dvs. tolking og behandling av forkortelser, akronymer, spesialtegn, valuta, tall og talluttrykk. I slike tekster er det mye som kan gå galt, slik som disse respondentene bekreftet:

Det store problemet med kunstig tale er at når man kommer inn på vanskelige temaer, får man ikke med seg innholdet. Det er feil lesing av forkortelser og vanskelige paragrafer. Det blir vanskelig. Men ren og enkel tekst uten noen forkortelser går greit. Særlig forkortelser i jus er umulig. Hadde det vært mulig å ha gjort slik at talesyntesen kunne lese jusforkortelser i jusbøker? For eksempel at al. leses aksjeloven? I justekster er det særlig vanskelig med uttrykk bestående av tall og tekst. (Svaksynt, 50 år)

Jeg har liten erfaring, men romertall I uttales riktig. Da er det rart at romertall II uttales feil. Jeg erfarer at det også er mye feil ved andre TTS-systemer når det gjelder tall og talluttrykk. (Annet handikap, 21 år)

Sakprosa med mye sammenhengende tekst uten altfor mange innslag av utenlandske navn, årstall og titler, fungerer bra med Brage. (Svaksynt, 64 år)

Alle disse tekstuelle utfordringene som ble kommentert her, er særlig krevende å tolke datamaskinelt. Som tabell 15 ovenfor angir, hadde særlig respondentene som studerte enten økonomi, administrasjon og ledelse eller jus, mye å utsette på Brages opplesing av slik fagterminologi. Økonomisk og juridisk studielitteratur vil kunne bestå av mange tekstuelle utfordringer. Reint syntaktisk kan også språket i juridiske tekster være tungt å formidle selv for en profesjonell innleser. Språkrådet har foreslått noen retningslinjer for å gjøre språket i juridiske tekster mer tilgjengelig (Klarspråk, 2010). Tekstutdraget nedenfor er hentet fra ei sentral pensumbok for privatrettslig grunnstudium, og illustrer hvor utfordrende juridiske tekster kan være:

Hviler det odelsrett på en eiendom, kan en odelsberettiget gå til løsningssak mot lengstlevende - «det forhindrer verken skiftelov eller arvelov», Falkanger s. 296. Den odelsberettigede loddeier kan dog ikke påberope seg sl. § 62 – den forutsetter at det skjer et skifte. (Lødrup, 1990, s. 346)

For en fagperson som er kjent med jusfaget, vil det være enklere å avkode og formidle innholdet bak syntaksen og de tekstuelle utfordringer. Men her er det derimot mye som kan gå galt for en talesyntese. Dersom substantivet *loddeier* ikke er registrert i uttaleleksikonet, kan autogeneratoren, som et resultat av en krevende syntaktisk oppbygging, kanskje foreslå at subjektet i det andre punktumet er *Den odelsberettigede og loddeier* er et sammensatt verbal

med *lodd + eier*. I verste fall kunne en tung syntaks ha forårsaket at det refleksive pronomenet *seg* kunne ha blitt tolket som preteritumsforma av verbet *å sige*, og dermed ha blitt uttalt deretter. Regelsettet som Filibuster er utrustet med for å sonde mellom et refleksivt pronomen og et verb i preteritum, er basert på sannsynlighetsberegning og avgjøres ut ifra konteksten. En utilgjengelig syntaks kan kunne resultere i morfologisk feilmerking, som igjen vil føre til ukorrekt uttale. For eksempel i dette tilfellet kunne man ha laget en regel som sa at dersom et verb som *påberope* opptrer foran den ortografiske forekomsten *seg*, vil disse to forekomstene sannsynligvis utgjøre et refleksivt verb, og skal uttales deretter.

Det er med andre ord visse typer studietekster som vil være mer krevende å prosessere enn andre. Om det ikke er nok at slike fagtekster i utgangspunktet kan være lite tilgjengelig for en student; dersom Brage leser tekstene med galt ordtrykk og tonelag på homografer, og feil gjengivelse av lovforkortelser og talluttrykk, vil tekstene bli enda mindre tilgjengelige.

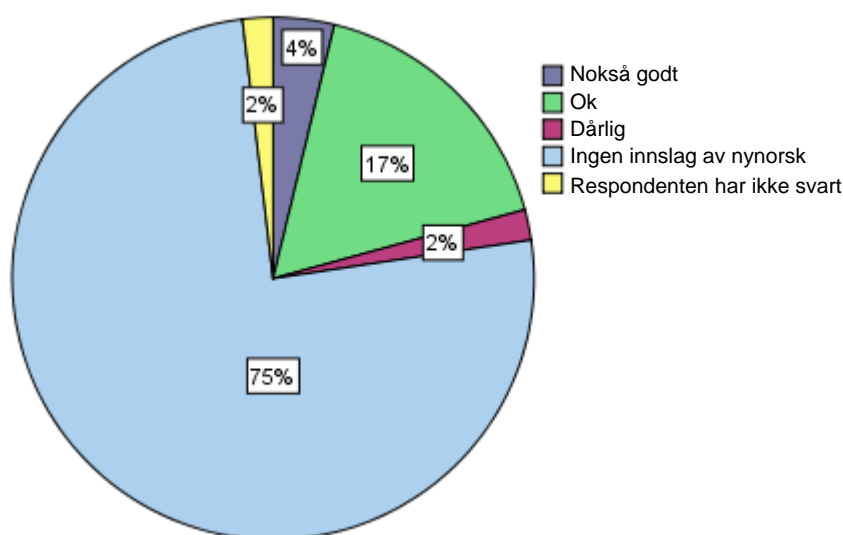
Men allikevel var det altså én av fire respondenter som studerte økonomi, som syntes at Brage mestret økonomisk fagterminologi på en god måte. Til forskjell fra de tre andre, som hadde benyttet seg av en norsk talesyntese i mindre enn ett år, oppga denne respondenten at han hadde benyttet seg av en norsk talesyntese i mer enn ti år, både i forbindelse med lesing, skriving og navigering på datamaskin, telefon og i forbindelse med studier. Og de to respondentene som studerte jus, som heller ikke hadde noen positiv opplevelse med Brages opplesing av juridisk fagterminologi, oppga at de hadde anvendt en norsk talesyntese mellom ett til tre år. Igjen kunne dette synes å gi en pekepinn på at vane og trening med syntetisk opplesing over tid vil kunne gi bedre forståelse (Hjelmquist et al., 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Men dette kan kun forklares ut ifra respondentenes erfaring med talesyntese på generelt grunnlag, ettersom dette ikke bare vil være utfordrende for Brage. Man kan ikke overse det faktum at slik type studielitteratur er krevende å gjengi tydelig med syntetisk opplesing.

Tolking og behandling av nynorsk

Selv om Brage først og fremst er utviklet for norsk bokmål, er talesyntesen utstyrt med et uttaleleksikon som inneholder enkelte nynorske posteringer. Uttaleleksikonet inneholder i dag (1. september 2010) 1.610 nynorske ordforekomster, som utgjør 0,2 % av hele leksikonet.

Egentlig vil antallet nynorske registreringer være langt mindre, siden dette tallet også inkluderer alle bøyningsformer av det samme ordet. Ved sida av for eksempel lemmaet *øyre* (substantiv, intetkjønn, entall, ubestemt form) er det registrert følgende morfologiske bøyningsformer: *øyret, øyrets, øyrer, øyrers, øyra, øyras, øyrene, øyrenes*. Dette er med andre ord bare et minimum for å sikre at de mest frekvente nynorske ordene uttales korrekt. Da er det kanskje nokså tankevekkende at Brage i dag også brukes til å produsere studielitteratur skrevet både på bokmål og nynorsk. Men siden så mange som tre av fire respondenter oppga at de ikke hadde noen innslag av nynorsk i studielitteraturen sin i det hele tatt, vil nok denne utfordringa med nynorsk nedprioriteres i den framtidige videreutviklinga.

Figur 6. Fordeling av utvalget etter hvordan studentene syntes Brage mestret innslag av nynorsk. Prosjenter. N=53.



Bemerk at ingen av respondentene syntes Brage leste nynorsk på en god måte. Derimot syntes 4 % at Brage mestret innslag av nynorsk nokså godt, 17 % på en ok måte, og 2 % på en dårlig måte. Den ene studenten som ikke oppga noe svar, følte han ikke hadde tilstrekkelig med erfaring til å kunne uttale seg.

Det må her også tas høyde for at opplesing av nynorsk på setningsnivå kan oppleves annerledes enn på ordnivå, noe som dette resultatet ikke sonderer mellom. Dersom Brage skal uttale ei hel nynorsk setning, må det altså tas høyde for prosodiske forhold som enkelte kanskje ville ha reagert på. Brage vil lese nynorsk med en østlandsk setningsmelodi. For å lese nynorsk med vestlandsk setningsmelodi, burde det ha blitt spilt inn en egen dedikert

taledatabase med dette formålet. Men for det første krever dette et eget nynorsk uttaleleksikon, og ikke minst et budsjett som kan forsvares både i forhold til kostnadene og tida det vil ta å gjennomføre dette. Så lenge Brage gjør seg forstått under opplesing av nynorske innslag, må dette være akseptabelt for en bokmåtalesyntese med østlandsk setningsmelodi.

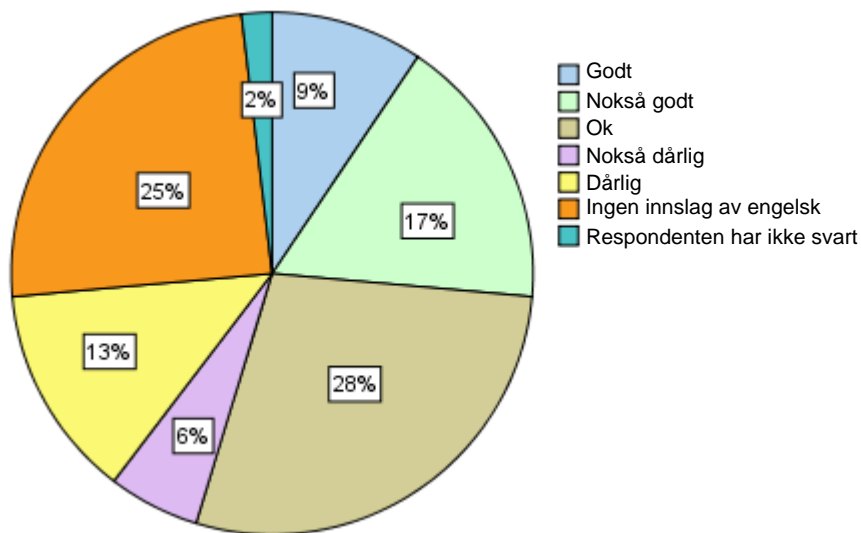
Tolking og behandling av engelsk

Før respondentene opplevelse av engelskferdighetene til Brage diskuteres, må disse ferdighetene settes i kontekst. Ei nyhetssak som ble vist i TV2 Nyhetene 7. september 2010, vil illustrere dette. Denne saka er også publisert på nettsidene til TV2. Her kan en lese at *en glipp hos Skatteetaten fikk en norsktalende datamaskin til å snakke engelsk – resultatet er hysterisk morsomt, og du kan høre det her* (Bentzen, 2010). Talesyntesen som Skatteetaten benyttet som oppleser på nettsidene sine var Acapela Groups Kari, en etablert og god likt norsk talesyntese. NLB produserte studielitteratur med denne fra 2004 til Brage overtok i slutten av 2009, og tre respondenter oppga at de fremdeles foretrakk Kari framfor Brage som oppleser av studielitteratur. Men som nyhetsjournalistene i TV2 observerte er Kari ikke hovedsakelig trenet til å lese engelsk. I stedet kan Acapela Group tilby hele ni høykvalitetssynteser for engelsk, slik at sluttbrukeren selv kan velge mellom en norsk og en engelsk talesyntese etter behov (Acapela Group, 2009).

Å bruke både en norsk og en engelsk talesyntese vil gi det beste resultatet. Men siden bibliotek som svenske TPB og norske NLB hadde behov for en talesyntese som særlig skulle tolke og behandle tekstuelle utfordringer som særpreger studielitteratur, hvor engelske innslag forekommer hyppig, ble det gjort et forsøk på å la en og samme talesyntese håndtere morsmålet (svensk eller norsk) i tillegg til engelsk. Dette må gjøres ut ifra realistiske forventninger, hvilket vil innebære at engelskuttalen vil få en fornorsket uttale og setningsmelodi. Under innspillinga av taledatabasen ble det lest inn i underkant av 1.150 engelske talesekvenser, som ble hentet fra manuskriptet til CMU ARCTIC-databasen (Kominek og Black, 2003). Dessuten ble det også lest inn sekvenser bestående av karakteristiske fonemer for tysk og fransk, slik at tyske og franske passasjer og egennavn skulle kunne uttales på en tilstrekkelig tydelig måte.

Med bakgrunn i dette, vil det være urimelig å kreve at Brage skal lese engelsk med perfekt uttale og prosodi. Tanken er først og fremst at engelsk skal være forståelig, og ikke nødvendigvis perfekt.

Figur 7. Fordeling av utvalget etter hvordan studentene syntes Brage mestret innslag av engelsk. Prosent. N=53.



En fjerdedel (26 %) av respondentene syntes Brage mestret engelsk enten godt eller nokså godt, og 28 % syntes Brage mestret engelsk på en ok måte. Men det var 19 % som opplevde at Brage leste engelsk enten dårlig eller nokså dårlig. Den resterende fjerdedelen av respondentene (25 %) oppga at de ikke hadde noen erfaring med syntetisk opplesing av engelsk i studielitteratur. Den ene respondenten (2 %) som ikke oppga noe svar, begrunnet dette med at han følte han ikke hadde tilstrekkelig erfaring til å uttale seg.

Dette må være et akseptabelt nivå av aksept for engelsk uttale, særlig med tanke på at Brage er en norsk bokmålstalesyntese.

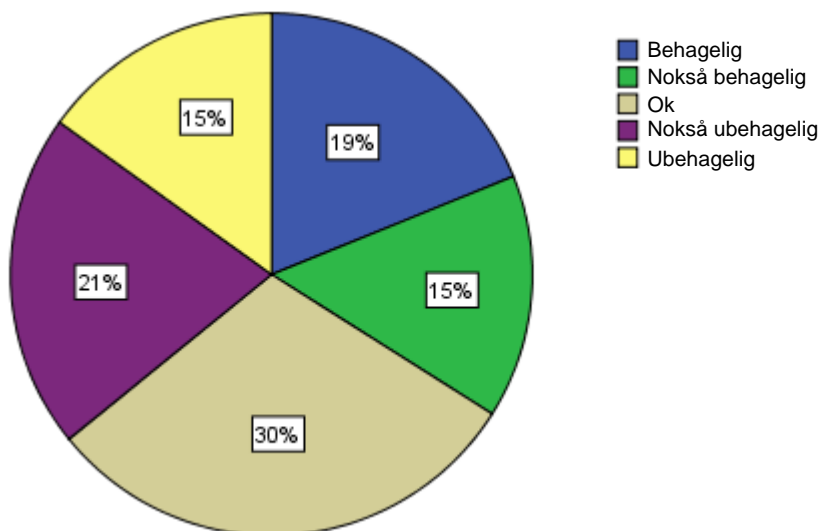
Formidlingsevne

Flere forskere anbefaler bruken av talesyntese i undervisning for å hjelpe elever med lesevansker, som refererer til studier som viser at et slikt hjelpemiddel vil bidra til å lette oppgava disse barnene står overfor med å avkode og forstå tekst (Olofsson, 1992; Wise,

Olson og Ring, 1997; Van Daal og Reitsma, 2000; Lewin, 2000). Men til tross for dette og det faktum at kanskje om lag 10 % av Norges befolkning har lese- og skrivevansker (Kamstrup et al., 2002, s. 2), oppga så mange som 30 respondenter (57 %) at de kun hadde benyttet en norsk talesyntese i mindre enn ett år, hvor 19 av disse 30 respondentene (63 %) hadde dysleksi, og to (4 %) respondenter at de ikke hadde benyttet talesyntese i det hele tatt. For oppsiktsvekkende mange var Brage deres første møte med en talesyntese.

Tekstformidling må i denne sammenhengen forstås som en talesynteses evne til å formidle en tekst til en lytter på en slik måte at meningsinnholdet i teksten blir forstått. Ut ifra ei slik oppfatning vil tekstformidling i sammenheng med talesyntese være et sammensatt begrep, som må defineres som summen av hvor naturlig og hvor tydelig respondentene opplevde Brage. Det overordnede evalueringskriteriet som talekvaliteten til Brage må måles etter i forhold til formidlingsevne, er kriteriet om velegnethet (Mariniak, 1993; Jekosch, 2005; King, 2007). I hvilken grad syntes respondentene at Brage løste den oppgava denne syntetiske stemmen i utgangspunktet var pålagt, nemlig å lese opp studielitteratur? Som sagt vil dette være påvirket av mange av de momentene som allerede er blitt diskutert, som for eksempel en naturlig prosodi, tydelig artikulasjon m.m.

Figur 8. Fordeling av utvalget etter hvordan respondentene opplevde det var å lytte til studielitteratur, som ble opplest av Brage, over tid. Prosentar. N=53.



34 % av respondentene syntes at det enten var behagelig eller nokså behagelig å lytte til studielitteratur, som ble lest opp av Brage, over tid, mens 30 % av respondentene syntes det

var ok, og de resterende 36 % respondentene syntes det enten var ubehagelig eller nokså ubehagelig. Fem av de åtte respondentene som syntes det var ubehagelig å lytte til Brage over tid, studerte samfunnsvitenskapelige studier, mens de resterende tre respondentene var fordelt etter hvilken studieretning de deltok i, slik: Jus (1), økonomi, administrasjon og ledelse (1) og annet (1). Tidligere ble det påpekt at juridisk og økonomisk faglitteratur er utfordrende for en talesyntese. Men hva kunne skyldes at seks av elleve studenter med samfunnsvitenskapelig bakgrunn enten syntes det var ubehagelig eller nokså ubehagelig å lytte til studie bøker med Brage over tid? Kunne det henge sammen med at Brage tolket samfunnsvitenskapelig fagterminologi på en dårlig eller nokså dårlig måte, slik som alle disse seks respondentene oppga? Uttaleleksikonet til Filibuster må vurderes ut ifra dekningsgraden av sentrale samfunnsvitenskapelige fagtermer og egennavn, siden samtlige av disse seks respondentene enten hadde delvis eller ingen tillit til Brage som formidler av studiepensumet deres. Denne vurderinga av uttaleleksikonet bør prioriteres i videreutviklinga.

Tabell 16. Fordeling av utvalget etter om respondentene opplevde at de måtte anstrenge seg mer for å holde konsentrasjonen oppe da de lyttet til DAISY-bøker, som ble lest opp av Brage, sammenliknet med DAISY-bøker med menneskelig opplesing. Resultatene er videre sammenliknet med hvor lenge studentene hadde benyttet seg av en norsk talesyntese. N=53.

ERFARING MED TALESYNTESE							
Anstrengelse	Mindre enn 1 år	Fra 1-3 år	Fra 4-6 år	Fra 7-9 år	10+ år	Ikke brukt TTS	Total
Måtte anstrenge seg mer	25	9	3	1	4	2	44
Måtte ikke anstrenge seg mer	1						1
Opplevde ingen forskjell	4		2	1	1		8
Total	30	9	5	2	5	2	53

Hele 83 % av studentene opplevde at de måtte anstrenge seg mer for å holde konsentrasjonen oppe med syntetisk opplesing enn med menneskelig. Dette funnet synes å samsvare med øvrige studier om talepersepsjon (Goldstein, 1995; Engstrand, 1993; Axmear et al., 2005). Men hvem var de øvrige ni studentene (17 %), som ikke følte at de behøvde å anstrenge seg mer ved syntetisk opplesing? Bare tre av disse ni studentene oppga at det var *viktig* for dem at syntetisk tale liknet menneskelig tale, mens de resterende fire respondentene oppga at det var *nokså viktig* og to respondenter at det verken var viktig eller uviktig:

Det er ikke noe problem. Har man nettopp hørt en menneskelig innlest bok, merker man forskjell. Men det er bare i starten. Etter hvert tenker man ikke på at det er en datastemme. Det går helt greit. (Dyslektisk, 22 år)

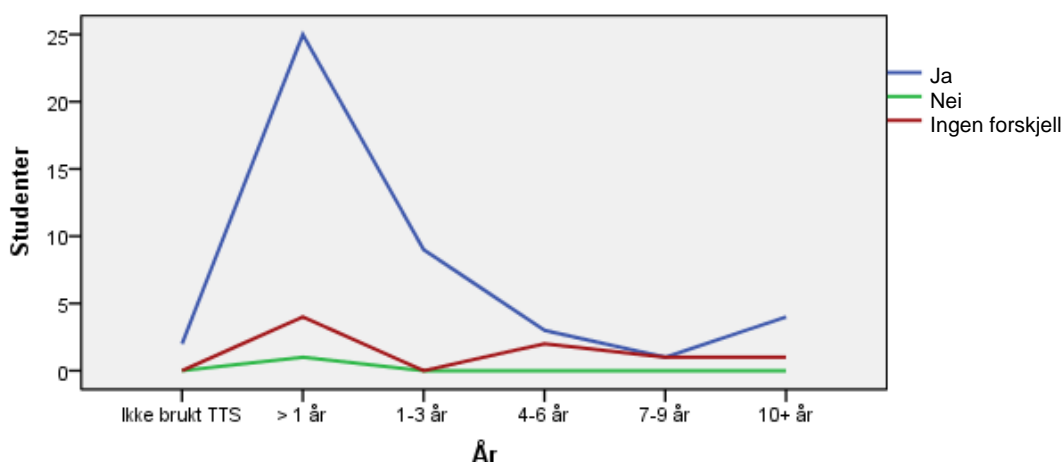
[Jeg] blir mye mindre sliten av å lytte til Brage sammenliknet med å lytte til Kari. Eller kanskje: Jeg sovner mye lettere av å lytte til Brage. Brage er en av de stemmene jeg blir mindre sliten av. (Svaksynt, 47 år)

Ikkje alltid behageleg å sitte å konsentrere seg over lengre tid. Dersom dei som les ikkje forstår kvad ei les, er vi like langt. (Dyslektisk, 56 år)

Studenten som oppga at hun etter hvert som hun kom ut i boka ikke merket forskjell på syntetisk og menneskelig opplesing, hørte trolig til unntaket. Men det var interessant at en student sammenliknet med en annen norsk talesyntese, og begrunnet svaret sitt ut ifra denne sammenlikninga. En tredje student sammenliknet med menneskelig opplesing og begrunnet svaret sitt med at selv ikke de profesjonelle innleserne alltid bidro til god formidling.

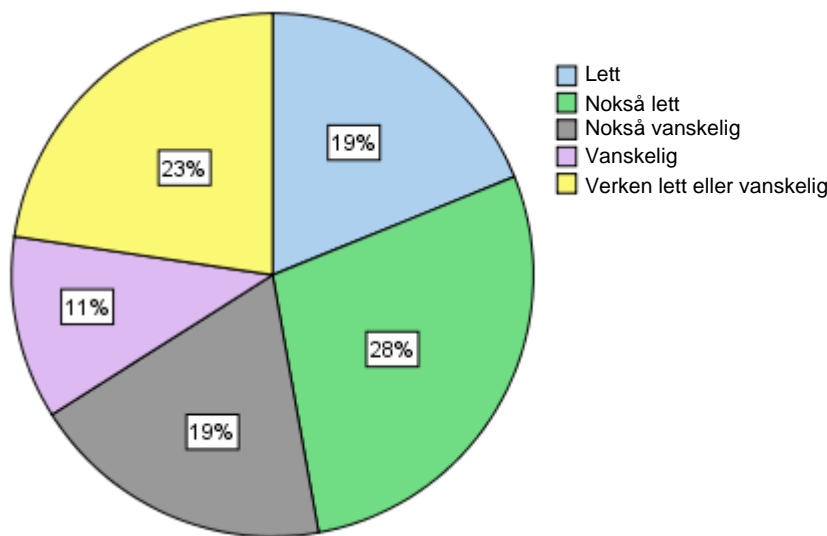
Nedenfor er dette resultatet sammenliknet med respondentenes erfaring med bruk av talesyntese. Den første grafen (blå) angir respondentene som opplevde at de måtte anstrenge seg mer for å holde konsentrasjonen oppe. Den andre grafen (grønn) angir dem som ikke opplevde at de måtte anstrenge seg mer. Den tredje grafen (rød) angir dem som ikke opplevde noen forskjell mellom syntetisk og menneskelig opplesing.

Figur 9. Fordeling av utvalget etter om studentene opplevde at de måtte konsentrere seg mer for å holde konsentrasjonen oppe da de lyttet til DAISY-bøker, som er produsert med Brage, sammenliknet med DAISY-bøker, som er produsert med menneskelig innlesing. Resultatene er videre sammenliknet med hvor lenge studentene hadde benyttet seg av en norsk talesyntese.



Det er den blå grafen som er interessant å studere nærmere. Den indikerer en sammenheng mellom konsentrasjon og talepersepsjon av syntetisk opplesing, og erfaring med bruk av talesyntese. De som hadde benyttet seg av en norsk talesyntese i mindre enn ett år, opplevde at de måtte konsentrere seg mer (Hjelmquist et al., 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007).

Figur 10. Fordeling av utvalget etter hvordan respondentene syntes det var å få med seg meningsinnholdet i en tekst, som ble opplest med Brage. Prosent. N=53.



47 % av respondentene syntes det var lett eller nokså lett å få med seg meningsinnholdet i en tekst, som ble lest opp av Brage, mens 30 % syntes det enten var vanskelig eller nokså vanskelig. Som tabell 17 nedenfor viser, syntes det å være en sammenheng mellom hvor lett det var å få med seg meningsinnholdet i en tekst, som ble opplest av Brage, og hvordan respondentene opplevde at Brage mestret fagtermer og -uttrykk innenfor deres studieretning.

Tabell 17. Fordeling av utvalget etter hvordan respondentene syntes det var å få med seg meningsinnholdet i en tekst, som ble lest opp av Brage (rader), sammenliknet med hvordan de syntes Brage mestret faguttrykk innenfor studieretninga si (kolonner). N=53.

MESTRING AV FAGTERMINOLOGI							
FORSTÅELSE		Godt	Nokså godt	Ok	Nokså dårlig	Dårlig	Total
	Lett	2	5	3			10
	Nokså lett	2	3	7	2	1	15
	Nokså vanskelig			5	3	2	10
	Vanskelig				3	3	6
	Verken lett eller vanskelig		1	6	3	1	12
	Total	4	9	21	11	7	53

Blant de 13 respondentene som syntes Brage mestret faguttrykk innenfor deres studieretning enten godt eller nokså godt, oppga 92 % av disse respondentene at det enten var lett eller nokså lett, og de resterende 8 % at det verken var lett eller vanskelig å få med seg meningsinnholdet. Og blant de 21 respondentene som syntes Brage mestret faguttrykk på en ok måte, syntes ingen at det var vanskelig, og fem respondenter at det var nokså vanskelig, å få med seg meningsinnholdet.

Dette funnet må forstås slik at videreutvikling av uttaleleksikonet er avgjørende for hvor velegnet Brage vil være til å lese studielitteratur. Som tidligere nevnt må dette arbeidet foretas ut ifra den gjeldende dekningsgraden av fagterminologi for hver enkelt studieretning.

Korrigerings av skrivefeil og skannefeil

Å forlange at en talesyntese skal være i stand til å korrigere uttale etter skrivefeil i den trykte teksten, og skannefeil, syntes urimelig. Dette er med andre ord feil som har oppstått under konverteringa av trykt til elektronisk tekst. Allikevel er dette utvilsomt et godt poeng. For de respondentene som poengterte dette, representerer slike feil allikevel en viss forringelse av den syntetiske talekvaliteten sammenliknet med menneskelig:

Når det er feil i den trykte boka er det vanskelig å bruke talesyntese, som ikke retter opp feil slik et menneske gjør, særlig i forbindelse med tegnsetting. (Annet handikap, 21 år)

Slike feil kan kanskje glemmes i ei måling av talekvaliteten. Men det er utvilsomt et poeng at den menneskelige evna til å rette skrive- og skannefeil under innlesing, ikke kan skapes i samme grad hos en talesyntese. Ei alternativ tilnærming til å unngå skrive- og skannefeil ved tilrettelegging og produksjon av studie bøker, kan være å søke etter disse feiltypene slik det er beskrevet i kapitlet bakerst som tar for seg ulike mulige løsninger for hvordan det videre utviklingsarbeidet skal tas, jf. Kapittel 7, s. 120.

Respondentenes preferanser og vaner

Tabell 18. Krysstabell som viser hvor lenge studentene hadde benyttet seg av en norsk talesyntese og hvilken talesyntese de foretrakk.

	ERFARING						Total
	Mindre enn 1 år	1-3 år	4-6 år	7-9 år	10+ år	Ikke benyttet TTS	
TALESYNTSE	Brage	9	2	2	1	1	15
	eSpeak norsk			1			1
	Kari	1	2				3
	Nora			1		1	2
	Stine	1	1			1	3
	Vegard			1		1	2
	andre	1				1	2
	Foretrekker ikke TTS	17	4		1	2	24
	Ubesvart	1					1
	Total	30	9	5	2	5	53

Det er særlig bemerkelsesverdig at hele 30 respondenter (57 %) hadde brukt en norsk talesyntese i mindre enn ett år. Blant disse var det 17 stykker som kun hadde hørt Brage, og ei som svarte at hun *syntes datastemmene var altfor vanskelig å følge [...] og derfor brukte [...] dem lite*. Dette innebærer at mange av respondentene hadde liten kjennskap til talesyntese, og for det meste bare hørt Brage som en del av tilbudet fra NLB. Nesten halvparten av respondentene (45 %) ønsket ikke å få studielitteratur tilrettelagt med talesyntese i det hele tatt. En slik motvillighet til å benytte seg av dette tilbudet kunne synes å være et resultat av

fordommer mot syntetisk tale generelt. Hjelmquist (et al., 1990) observerte at blinde og svaksynte som ble bedt om å lese aviser med syntetisk tale, gradvis vente seg til den syntetiske stemmen, og over tid beskrev talekvaliteten som god. Flere studier viser tilsvarende korrelasjon mellom vane og bedømmelse av syntetisk talekvalitet (Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Da så mange som 17 av de 24 studentene (71 %), som ikke foretrakk å få studielitteratur tilrettelagt med syntetisk tale, hadde brukt en norsk talesyntese i mindre enn ett år, syntes det å være en sammenheng mellom vane og aksept blant studentene.

Enkelte studenter, som hadde benyttet seg av en norsk talesyntese over lengre tid, nevnte også at vane spiller ei vesentlig rolle for brukeropplevelsen, som for eksempel studenten som fremdeles foretrakk Vegard etter ti år. Dette til tross for at Vegard er en forrige generasjons talesyntese, som enkelte vil hevde høres mindre naturlig ut enn skjøtesyntesene med enhetsutvelgelse som representerer dagens teknologi. Dette kan også være årsaka til at ingen av de blinde studentene foretrakk Brage. De blinde studentene bruker talesyntese, men riktignok andre talesynteser enn Brage, i langt større grad enn for eksempel dyslektikerne for å kunne lese og orientere seg både på datamaskinen, for eksempel i forhold til å lese og skrive e-post og surfe på Internett, og på mobiltelefon, for å lese og skrive tekstmeldinger og navigere seg gjennom menyen.

Besvarelsene nedenfor antyder at vane synes å spille ei avgjørende rolle for hva en foretrekker:

Noe har vært vanestyrkt. Den talen du har brukt på PC er av gammel vane. Man holder seg til den man er kjent med. (Blind, 42 år)

Det har vært mye pga. at dette er de talesyntesene som ligger inne i skjermleser og andre TTS-programmer. Begynte med talesyntese på PC, før på lydbøker. Tenker ikke over hvem man bruker når jeg sitter og skriver. Bruker den jeg er vant med. (Svaksynt, 31 år)

Dragvoll bibliotek benytter Nora i forbindelse med skanning av artikler. (Dyslektisk, 33 år)

Men sammenliknet med de andre norske talesyntesene kommer Brage meget godt ut; tre studenter foretrakk Kari og Stine kontra 15 som foretrakk Brage. Det må her tas høyde for at fem respondenter oppga at de heller ikke hadde benyttet seg av andre norske talesynteser enn Brage. Disse respondentene vil ikke ha noe særlig sammenlikningsgrunnlag å velge ut ifra. Med bakgrunn i dette bør dette tallet heller forstås som en aksept for Brage og talesyntese generelt. Tilsvarende oppga seks respondenter at de bare hadde brukt Brage og

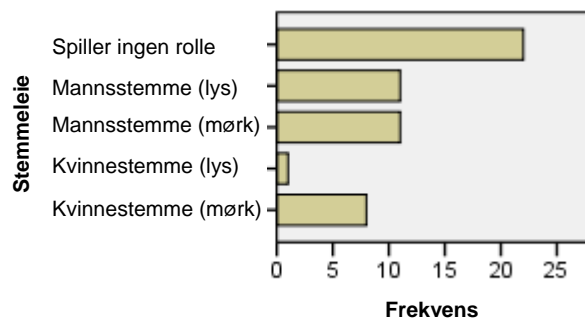
Kari, hvilket muligens kan forstås slik at de kun hadde kjennskap til de stemmene som de ble presentert overfor av NLB. Riktignok var det interessant at så mange som ni av disse femten studentene kun hadde brukt norsk talesyntese i mindre enn ett år. Og blant disse ni var det fire studenter som oppga at de hadde kjennskap til andre norske talesynteser. Dette antyder at disse studentene har hatt generelt gode brukeropplevelser med Brage, selv om de hadde liten kjennskap til andre norske talesynteser.

Kjønn og stemmeleie

Kvinnelige syntetiske stemmer har først blitt vanlig i de seinere årene. Før 90-årene syntes den generelle talekvaliteten blant de kvinnelige stemmene å framstå som mindre naturlig sammenliknet med de mannlige stemmene (Henton, 1992). Woods (referert til i Henton, 1992, s. 740) gjennomførte en studie som antyder at det er større variasjon i setningsmelodien hos engelsktalende kvinner sammenliknet med hos menn, som kan forklare hvor kvinnelige talesynteser er mindre naturlige enn mannlige. Klatt og Klatt (1989, s. 825) viser til ei rekke anatomiske forskjeller som ifølge dem resulterer i varierende talekvalitet hos en kvinnelig og mannelig leser. Så seint som i 2009 understreket Simpson (2009, s. 1) at aspirasjonsstøy fremdeles var ei generell utfordring for kvinnelig talekvalitet.

Men til tross for denne bakgrunnen oppga 22 av respondentene (42 %) at det ikke spilte noen rolle hvilket kjønn eller stemmeleie en talesyntese hadde. Dette synes å samsvare med det Campbell (2007, s. 34) hevdet om at vi ubevisst overse prosodiske egenskaper som hvert kjønn representerer. Det må her også tas høyde for at bare 13 av disse studentene oppga at de hadde benyttet seg av en kvinnelig syntetisk stemme, slik at de resterende ni studentene, som det ikke spilte noen rolle for, heller ikke hadde noe sammenlikningsgrunnlag.

Figur 11. Fordeling av utvalget etter hvilket kjønn og stemmeleie som var å foretrekke til en syntetisk stemme. Frekvens. N=53.



Blant de respondentene som valg av kjønn og stemmeleie spilte ei rolle for, syntes flere å foretrekke en mannsstemme (42 %) enn en kvinnestemme (16 %). Bare én respondent foretrakk å få opplest studielitteraturen med en lys kvinnestemme. Kommentarene som er gjengitt nedenfor gjenspeiler ytterpunktene i besvarelsene:

De kvinnelige har dårligere gjengivelse fordi de er såpass skarpe. Det er bedre med en mørk mannsstemme. Kvinnestemmer er vondt å høre på i lengden. (Dyslektisk, 43 år)

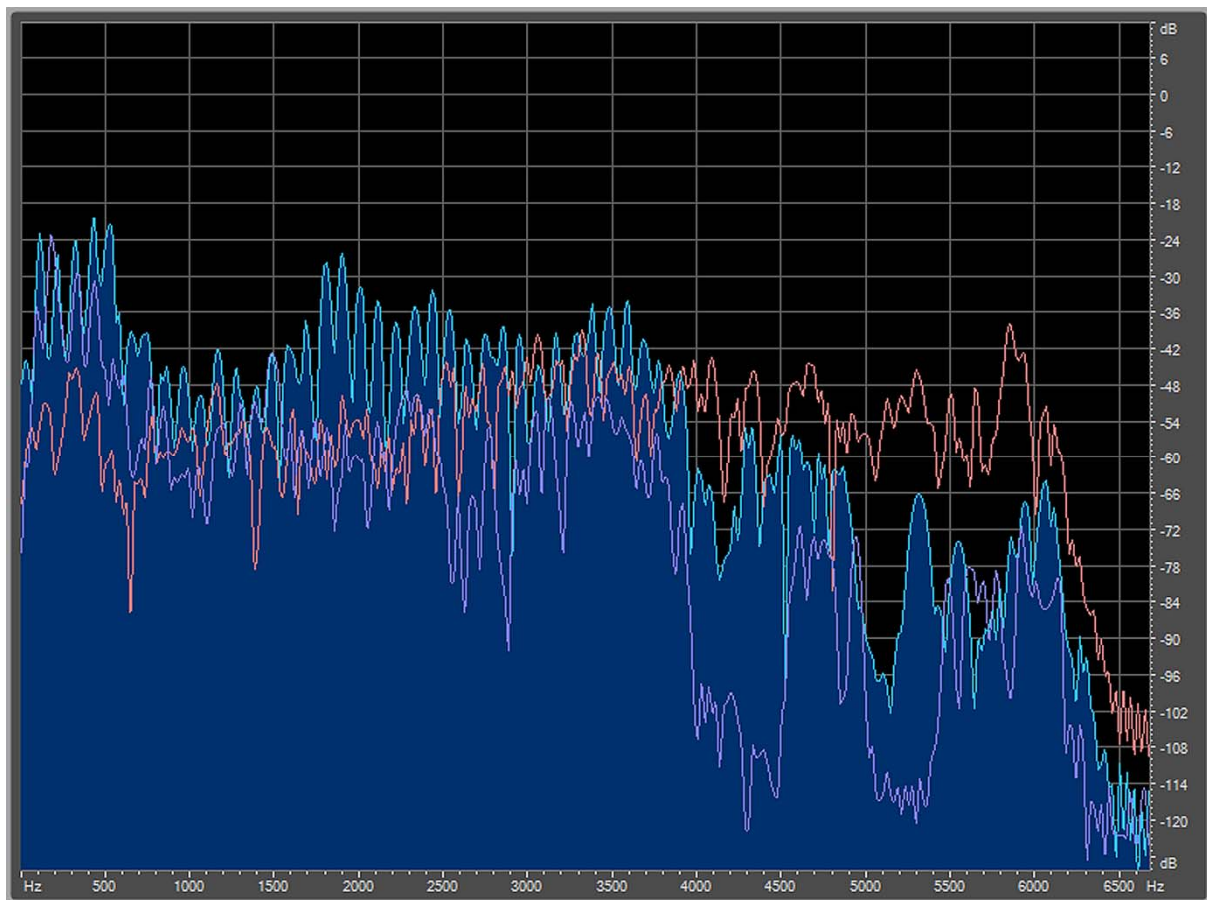
Brage har en mørk stemme. Jeg har så mye sus i ørene, slik at en lys stemme vil bryte med den [susen]. En mørk stemme toner den ned, mens en lys stemme forsterker tinituslyden. (Dyslektisk, 46 år)

Så mørkt at det er vanskelig å oppfatte på DAISY-spillere. (Blind, 27 år)

Litt vel mørkt. (Dyslektisk, 33 år)

For å kommentere den blinde studenten på 27 år, som hadde vanskeligheter med å oppfatte Brage på DAISY-spillere sin, er det nedenfor gjengitt et utdrag fra frekvensanalysen av stemmen til ”mannen bak Brage” (Olaussen og Haugen, 2008, s. 11-17).

Figur 12. Utdrag fra frekvensanalysen av stemmen til mannen bak den syntetiske stemmen Brage.



Figur 12 viser måling av stemmekvalitative egenskaper som volum og bæreevne, tydelighet, tilstedeværelse og formidlingsevne. I koordinatsystemet er frekvens (målt i Hz) angitt langs x-aksen, og volum (mål i dB) angitt langs y-aksen. Fargekodene beskriver stemmekvalitative egenskaper ved inngangen (turkis), midten (oransje) og ved utgangen (purpur) av ei tilfeldig framstilt setning. Kort beskrevet viser analysen at stemmen til Brage har god bæreevne ut hele setninga. I tillegg til å gi veldefinerte utslag i områder som gir stemmen attributter som fylde og tydelighet, uteblir også uheldige karakteristiske egenskaper, som for eksempel visslende s-lyd (Olaussen og Haugen, 2008, s. 11-17).

Selv om det kun var én respondent som opplevde at stemmen til Brage var så mørk at det var vanskelig å få den tydelig gjengitt, var det likevel et reelt problem for den studenten det gjaldt. Da frekvensanalysen antyder det motsatte kunne det tenkes at dette kanskje var forårsaket av andre faktorer, som for eksempel at høyttalere eller hodetelefoner med ukorrekt lydgjengivelse eller elementskade.

Grad av tillit til Brage som oppleser av studielitteratur

Graden av tillit til Brage som oppleser av pensumlitteratur må ses i sammenheng med det overordnede evalueringskriteriet om velegnethet (Jekosch, 2005). Ei slik måling vil kunne fortelle noe om hvilke holdninger de hadde til Brage og talesyntese generelt. Det er gjort ei rekke studier hvor det er forsøkt å måle holdninger og aksept i forhold til bruk av talesyntese (Léwy og Hornstein, 1994; Keller, 2000; Cryer og Home, 2009). Så vidt meg bekjent er det ikke gjennomført noen studier om holdninger og aksept til talesyntese i Norge, men i Sverige er det gjennomført flere studier hvor blant anna aksept i forhold til ulike svenske talesynteser ble forsøkt kartlagt (Hampshire et al., 1982; Fahlstedt, 2004; Persson, 2004).

Dette resultatet må ses i sammenheng med hvilken studieretning respondentene studerte innenfor. Å måle graden av tillit til Brage er en viktig målestokk for å si noe om den generelle talekvaliteten til Filibuster med Brage. Ifølge King (2007, s. 129-130) er ikke en talesyntese bedre enn det brukerne opplever at den er.

Tabell 19. Fordeling av utvalget etter hvilken studieretning respondentene studerte innenfor (rader), sammenliknet med hvilken grad av tillit de hadde til Brage som formidler av studielitteraturen deres (kolonner). N=53.

		GRAD AV TILLIT				
STUDIERETNING		Stor tillit	Delvis tillit	Liten tillit	Ingen tillit	Total
	Ex. Phil./Ex. Phac.	1				1
	Humanistiske fag	4	1	2		7
	Jus			2		2
	Matematikk og naturvitenskap	1	2	1		4
	Medisin	4	3			7
	Samfunnsvitenskap	5	2		4	11
	Tverrfaglig		1			1
	Utdanningsvitenskap	8	6			14
	Økonomi, administrasjon og ledelse	1		1	2	4
	Annet				1	1
	Videreutdanning			1		1
	Total	24	15	7	7	53

24 respondenter (45 %) hadde *stor* tillit til Brage som formidler av studielitteraturen deres, og 15 respondenter (28 %) hadde *delvis* tillit. De resterende 14 respondentene (26 %) hadde liten eller ingen tillit. Blant de sju respondentene som ikke hadde noen tillit i det hele tatt, syntes to respondenter at selve stemmen til Brage var velegnet, og to respondenter verken velegnet eller uegnet, til å formidle studielitteratur.

Men hva kunne ei slik manglende tillit være et resultat av? For det første kunne erfaring med bruk av norske talesynteser synes å spille ei viss rolle. Av de sju respondentene uten tillit til Brage, hadde fire respondenter benyttet en norsk talesyntese i mindre enn ett år, og to respondenter ikke i det hele tatt. Men riktignok hadde én respondent også benyttet seg av norske talesynteser i mer enn ti år. Denne respondenten oppga også at hun foretrakk Vegard framfor andre norske talesynteser, mens de resterende seks respondentene ikke foretrakk syntetisk tale i det hele tatt. Dette må forstås slik at disse respondentene syntes å ha en generell negativ holdning til talesyntese. Kommentarene nedenfor antyder at respondentene med ei manglende tillit vurderte hvor velegnet Brage var til å lese opp studielitteratur ut ifra flere sentrale evalueringskriterier (Mariniak, 1993; King, 2007):

Jeg synes talesyntese er en utrolig ubehagelig måte å lese på. Foretrekker da heller dårligere innlesere. (Svaksynt, 48 år)

Jeg bruker mer energi for å forstå ordene og sammenhengen i setningene, og for å tyde hvilket ord de faktisk sier. Mange som har dysleksi bruker øyne og ører mer enn andre når de leser, og når det tukles med informasjonen man får gjennom ørene (med tonefall og i det hele tatt), blir det vanskelig. (Dyslektisk, 26 år)

Vanskelig og enkelte ganger umulig å få med seg innholdet i pensum når det er innlest med talesyntese. Ingen setningsoppbygning og lite flyt i innlesingen. Jeg hoppet over bøkene som var innlest med talesyntese. (Dyslektisk, 32 år)

De to første kommentarene kunne synes å være begrunnet i Brages svake velegnethet i forhold til kriteriet om behagelighet, mens den tredje kommentaren særlig berørte kriteriene om tydelighet og naturlighet. Dette må forstås slik at siden disse respondentene opplevde talekvaliteten til Brage som lite tydelig, unaturlig og ubehagelig, var den dermed også ikke egnet til å formidle studielitteratur.

Videre kunne dekningsgraden av fagterminologi innenfor enkelte studieretninger være ei annen forklaring. Som tabell 19 ovenfor viser, oppga én student som tok forberedende emner, at hun hadde stor tillit til Brage, mens fem av sju studenter (71 %) som studerte

humanistiske fag hadde enten stor eller delvis tillit. Og tilsvarende hadde tre av fire respondenter (75 %) som studerte matematiske og naturvitenskapelige fag, sju av elleve respondenter (64 %) som studerte samfunnsvitenskapelige fag, og samtlige respondenter som studerte medisinske, tverrfaglig og utdanningsvitenskapelige fag, stor eller delvis tillit til Brage. Derimot syntes særlig respondenter innenfor studieretningene jus, økonomi, administrasjon og ledelse, og annet, å ha liten tillit til Brage. Dette innebærer at dersom Brage ikke kan formidle pensumbøker på en forståelig måte, syntes respondentene dette gjaldt heller ikke at talesyntesen var velegnet til å lese studielitteratur. Det er tidligere beskrevet hvilke utfordringer som særlig synes å gjelde for juridisk og økonomisk faglitteratur. Dette er tydelige domener Brage må trenes oppimot, både med tanke på repeterende syntaktiske utfordringer og med tanke på å tilføye manglende fagterminologi i uttaleleksikonet.

5.1.2 SUS-test

Neste del av denne studien omfattet ei objektiv måling av hvor tydelig Brage faktisk var, særlig på setningsnivå. Jeg har også vurdert datamaterialet på fonemnivå. I dette kapitlet vil resultatene fra SUS-testen bli framstilt og diskutert i lys av evalueringskriteriet om tydelighet.

Poengfordelinga mellom de tre talesyntesene

Gjennomsnittlig skåret Stine dårligere enn Kari og Brage. Av totalt 408 mulige rette, oppnådde Stine et resultat på 330 poeng. Dette var 33 poeng (8,09 %) mindre enn Kari og 56 poeng (13,73 %) mindre enn Brage. Ut ifra denne SUS-testen syntes Brage å være tydeligere enn Kari og Stine. I brukerundersøkelsen foretrakk hele 15 av 53 studenter Brage som oppleser til studielitteratur, mens bare tre studenter ønsket Stine og tre studenter Kari. Blant informantene i SUS-testen var det enkelte som nevnte at lesehastigheten påvirket hvor tydelig de opplevde de tre forskjellige talesyntesene. Det vil være vanskelig å måle graden av hvordan lesetempo påvirker tydelighet ved hjelp av en SUS-test. Men i stedet kan det være mulig å se om det var enkelte segmenter, segmentgrupper eller ord som gikk igjen, og om det var noen av disse tilfellene som syntes å være gjeldende for alle de tre talesyntesene.

Tabell 20. Resultatene fra SUS-testen fordelt som gjennomsnittsskår per informant og totalskår for seks informanter.

Talesyntese	Gjennomsnitt pr. informant	Gjennomsnitt pr. informant %	Totalt pr. 6 informanter
Brage	386	94,61 %	2.324
Kari	363	88,97 %	2.178
Stine	330	80,88 %	1.979

Man skal ta i betraktning at vane og erfaring med bruk av talesyntese kan være en faktor som vil påvirke tydelighetstester som denne SUS-testen, særlig i forhold til erfaring med å lytte til syntetisk tale i høy lesehastighet. Enkelte studenter jeg har vært i dialog med fortalte at de kunne lese med talesyntese i en hastighet som for personer uten særlig erfaring vil høres ut som ”grøt” (Hjelmquist et al., 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Med bakgrunn i dette kan det være at Stine ville ha skåret høyere i samme SUS-test dersom utvalget besto av informanter som hadde god erfaring med å lese raskt med talesyntese. Informantene til SUS-testen kommenterte også at særlig Stine var vanskelig å tyde siden hun leste så fort. Her hadde Brage en fordel framfor både Kari og Stine, ettersom denne stemmen har en vesentlig lavere lesehastighet, jf. Kapittel 5.1.1, s. 59-63.

Som sagt besto SUS-testen av fem forskjellige strukturer. Var det enkelte syntaktiske strukturer som var vanskeligere å tyde enn andre? Nedenfor er resultatene fordelt mellom de forskjellige strukturene. Resultatene viser gjennomsnittet av maksimumsskåren for hver setning innenfor ei strukturgruppe. For eksempel var det mulig å skåre seks poeng per setning innenfor den transitive strukturen. Ei setning ble lyttet til av seks forskjellige informanter per talesyntese, slik at det var mulig å skåre 36 poeng totalt for hver setning innenfor denne strukturgruppa. Som tabell 21 viser, skåret Brage 35 av 36 mulige poeng innenfor den transitive strukturgruppa. Hensikten med denne grundige statistiske framstillinga var å se om enkelte syntaktiske strukturer var mer utfordrende enn andre, og forsøke å finne ut hva dette eventuelt var forårsaket av. I tillegg var det ønskelig å kartlegge spredninga i poengfordelinga.

Tabell 21. Resultatene fra SUS-testen fordelt på hver av de fem syntaktiske strukturene.

		BRAGE	KARI	STINE
1. Intransitiv struktur	Gjennomsnitt	39 av 42	39 av 42	36 av 42
	Modus	42	39	36
	Standardavvik	4	3	4
	Variasjonsbredde	11	9	12
	Prosentil .25	37	37	33
	Median	41	39	36
	Prosentil .75	42	41	38
2. Transitiv struktur	Gjennomsnitt	35 av 36	31 av 36	30 av 36
	Modus	36	33	33
	Standardavvik	1	5	5
	Variasjonsbredde	4	17	16
	Prosentil .25	35	30	27
	Median	36	33	33
	Prosentil .75	36	35	35
3. Imperativ struktur	Gjennomsnitt	34 av 36	31 av 36	24 av 36
	Modus	36	32	22
	Standardavvik	3	4	5
	Variasjonsbredde	9	10	17
	Prosentil .25	33	28	20
	Median	35	32	23
	Prosentil .75	36	34	28
4. Spørrestruktur	Gjennomsnitt	45 av 48	43 av 48	39 av 48
	Modus	45	44	46
	Standardavvik	2	5	10
	Variasjonsbredde	6	15	28
	Prosentil .25	45	40	33
	Median	46	45	43
	Prosentil .75	47	47	47
5. Relativ struktur	Gjennomsnitt	40 av 42	38 av 42	37 av 42
	Modus	42	42	39

Standardavvik	4	5	5
Variasjonsbredde	10	14	14
Prosentil .25	39	33	33
Median	42	40	39
Prosentil .75	42	42	40

Disse statistiske data er inkludert for å kunne gi ei bedre framstilling av hvordan resultatene var fordelt mellom de tre talesyntesene enn bare å sammenlikne dem ut ifra gjennomsnitt og standardavvik. For det første skåret Kari og Stine gjennomsnittlig lavest på setninger med en imperativ struktur, som for eksempel på setninga *Tegn en kveld og en post*, hvor verbalet *Tegn* ble gjengitt feil av fire av seks informanter både hos Brage og Kari, og av fem av seks informanter hos Stine. Informantene hadde store vanskeligheter med å tyde setninger med imperativ struktur hos Stine. Denne strukturgruppa dro det totale resultatet betraktelig ned hos Stine. Poengene var gjennomsnittlig jevnt fordelt mellom de fem strukturgruppene hos Kari, men de intransitive setningene syntes å være noe tydeligere sammenliknet med de andre setningsstrukturene. Tilsvarende var poengene gjennomsnittlig jevnt fordelt mellom gruppene hos Brage, men de transitive setningene skilte seg positivt ut med at hele åtte av de tolv transitive setningene ble gjengitt 100 % ordrett av samtlige informanter. Lavest gjennomsnittsskår per setning hos Brage utgjorde de intransitive setningene. Men til de intransitive setningene kan vi se av prosentilene at fordelinga av skårene ligger jevnt over den prosentile fordelinga til både Kari og Stine.

For det andre varierer standardavvikene mellom de tre stemmene. Avvikene er for det meste nokså jevne, med unntak av gruppa med spørresetninger. Standardavviket til denne gruppa hos Stine er ti poeng kontra fem poeng hos Kari og to poeng hos Brage. Variasjonsbredda er nokså ujevn både i forhold til hver gruppe og mellom hver stemme. Størst forskjell mellom de tre stemmene var det ved spørresetningene fra seks poeng hos Brage, 16 poeng hos Kari, helt til 28 poeng hos Stine. Avslutningsvis må den høye skåren ved 2. prosentilen (0.25) hos Brage trekkes fram. Høy skår ved samtlige av de fem gruppene antyder at poengene var godt representert over gjennomsnittet. Videre antyder dette at det syntes å være mindre spredning i poengfordelinga sammenliknet med tilfellet hos Stine. Dette må leses i sammenheng med den høye skåren for medianen i hver enkelt gruppe.

Brage skåret med andre ord svært godt på alle de fem syntaktiske strukturene sammenliknet med Kari og Stine. Særlig sistnevnte hadde en vesentlig stor ustabilitet i poengfordelinga innenfor hver strukturgruppe. Av alle de 60 setningene ble sju setninger gjengitt helt ordrett med Stine, ti setninger med Kari, og 26 setninger med Brage. Kanskje enkelte vil synes dette var en lav totalskår, men tatt i betraktning at setningene ikke var meningsbærende må 26 av 60 setninger med Brage anses som et tilfredsstillende resultat. Og særlig dersom man tar i betraktning at poengfordelinga hos Brage var jevnt og stabilt fordelt med generelt høy skår ved alle gruppenes 2. prosentiler, gir dette et godt helhetsinntrykk for talepersepsjon på setningsnivå.

Kravet til SUS-testen om at informantene kunne høre ei setning kun én gang, kan vurderes i sammenheng med talepersepsjon. Dette vil ikke samsvare med hvordan NLBs lånere leser i virkelighet. Dersom en student kommer over et utydelig ord, vil det være naturlig å gå tilbake for å høre ordet igjen. Dette vil bli gjentatt til studenten enten er i stand til å gjette seg fram til det korrekte ordet eller må fastslå at ordet er helt uforståelig. Neovius og Raghavendra gjennomførte en studie i 1993 som bekreftet at gjentakende lyttinger resulterte i bedre forståelse og færre uforståelige ord (Neovius og Raghavendra, 1993, s. 21).

Et anna vurderingsmoment ved SUS-testen er at selv om SAM-gruppa understrekte at testen ikke skulle bli noen korttidshukommelsestest, hendte det gjentatte ganger under datainnsamlinga at informantene hadde problemer med å huske ei hel setning (Benôit, Grice og Hazan, 1996, s. 381-392). Ifølge Benôit, Grice og Hazan skulle korte setninger med maksimalt åtte ord forhindre nettopp dette. Men selv om det stort sett kun ble brukt enstavelsesord i den grad det var mulig, fikk enkelte av informantene vansker med å huske deler av setninga. Dette gjaldt særlig i de tilfellene hvor de måtte fokusere på ett eller flere helt uforståelige ord. Setninger uten altfor mange utydelige ord gikk tilsynelatende greit å gjengi uten at det gikk på korttidshukommelsen løs. Det er uvisst hva det kunne være forårsaket av. Men for å unngå at SUS-testen skulle utfordre korttidshukommelsen til informantene sørget jeg for å skrive ned mest mulig av det som ble sagt med en gang, slik at informantene ikke behøvde å huske hele setninga i tillegg til å dvele ved utydelige ord.

De fem minst tydelige setningene hos hver enkelt talesyntese

I tabell 22 er de fem minst korrekt gjengitte setningene til hver av de tre talesyntesene, framstilt. De er rangert på en slik måte at setninga med lavest skår er listet øverst.

Representerte disse setningene noe felles mellom de tre syntetiske stemmene?

Tabell 22. De fem minst tydelige setningene hos hver av de tre talesyntesene.

BRAGE	KARI	STINE
1. Et barn ba mot et trygt gram	1. En trygg tekst endte et fag	1. Når ville en ånd leie en lik bønn?
2. En rett ansatte en drift som tok	2. Når kunne en lov omfatte en rett slutt?	2. Når ville en lek glede en trygg feil?
3. Et gull eide en hud som etablerte	3. Et brev brukte en ro som økte	3. Del en stat og en tråd!
4. Varm et rom og en plan!	4. Når ville en lek glede en trygg feil?	4. Press et svar og en ferd!
5. Når ville et skilt hente en redd arv?	5. Et gull eide en hud som etablerte	5. Vis en rett og et håp!

Svaret er både ja og nei. For det første syntes de syntaktiske strukturene å være temmelig jevnt fordelt hos Brage og Kari. Med andre ord ga datamaterialet inntrykk av at det ikke var noen bestemt setningsstruktur som utfordret talekvaliteten mer enn andre. Det samme kan ikke sies å være tilfellet for Stine hvor kun to av de fem syntaktiske strukturene var representerte blant disse fem minst korrekt gjengitte setningene: den imperative strukturen og spørrestrukturen. Dette bekrefter funnet om at de imperative setningene skilte seg negativt ut i forhold til de andre strukturene hos Stine. Og i forhold til spørrestrukturen, hvor det riktignok også bør nevnes at typetallet til denne gruppa (46 poeng) var noe høyere enn hos både Brage (45 poeng) og Kari (44 poeng), kan en lese av tabell 21 at spørrestrukturen var den syntaktiske strukturen med størst variasjonsbredde hos Stine (28 poeng kontra seks hos Brage og 15 hos Kari). Ei så stor variasjonsbredde indikerer at talepersepsjonen av slike setninger med Stine var svært ujevnt fordelt mellom lytterne.

Det syntes å være en tendens til at verbalet i de imperative setningene ble slått sammen med den etterfølgende artikkelen, og ble dermed enten gjengitt som det samme verbet, men med endret tempus og/eller modus, eller som et substantiv. Dette vil bli drøftet seinere, jf. Kapittel 5.1.2, s. 101-102.

Feil ordklassifisering av homografer i testmaterialet

Det var ei rekke tilfeller hvor det allerede i testmaterialet forelå åpenbare feil, altså feil som ble generert fra systemene som de tre forskjellige stemmene var en del av. Disse feilene resulterte i at setningene hvor feilene forekom naturlig nok var vanskeligere for informantene å tyde sammenliknet med andre setninger. Det var særlig setningene med ble generert med Kari, som var preget av slike feil. For eksempel i setninga *En stein svarte en kveld som trengte* hadde antakeligvis tekstforprosessoren til Kari tolket verbet *svarte* som et adjektiv. Regelen i tekstforprosessoren som regnet ut sannsynligheten for at en homograf som dette enten skal være verb eller adjektiv syntes i dette tilfellet å være svak hos Kari. Denne homografen skal uttales med en kort /A/-lyd dersom det var et adjektiv, og lang /A:/-lyd dersom det var et verb. Siden jeg ikke har tilgang til systemet som Kari er en del av, kan jeg ikke med sikkerhet hevde at denne feilen er relatert til tekstforprosessoren. Alternativt kan dette være forårsaket av at uttalen ble konkatenerert med et galt lydsegment (/A/-lydsegmentet). Allikevel klarte to av seks informanter som lyttet til denne setninga med Kari å gjette seg fram til det riktige ordet. Sammenliknet med de to andre talesyntesene gjettet fire av seks informanter, som lyttet til samme setning med Stine, riktig uttale, mens alle som lyttet til setninga med Brage gjenga hele setninga korrekt.

Denne type feil, som etter vurderinga mi syntes å være forårsaket av gal ordklasseoppmerking, syntes å gå igjen hos Kari. Det kunne synes som om tekstforprosessoren til Kari ikke var trenet tilstrekkelig til å kunne skille homografer ut ifra konteksten. Et anna eksempel, hvor den samme homografen (*svarte*) ble gitt ukorrekt uttale, var i setninga *En fast sak svarte et hjem*. Igjen ble dette verbet antakeligvis klassifisert som et adjektiv. Men til forskjell fra forrige eksempel var det ingen av informantene som klarte å gjette seg fram til korrekt uttale. Derimot kunne fem av seks informanter, som lyttet til samme setning med Stine, gjengi hele setninga. Alle som lyttet til denne setninga med Brage kunne gjengi hele setninga korrekt.

Et anna tilfelle hos Kari der tekstforprosessoren syntes å ha merket en homograf med gal morfologisk informasjon, forekom i setninga *En kveld bidro uten en hel død*. Her leses *død* med en stum /d/ (i kodaposisjon), slik som om man ville ha uttalt adjektivet *død* i for eksempel setninga *En død katt lå i hagen*. Dette syntes å være en feil som var vanskeligere å tyde. I dette tilfellet klarte bare én av seks informanter å gjette seg fram til den riktige uttalen.

Alle informantene som lyttet til samme setning med Stine kunne gjengi hele setninga korrekt, mens fem av seks informanter som lyttet med Brage kunne gjengi samme setning korrekt.

Ingen tilfeller av homonymirelaterte feil ble observert verken hos Stine eller Brage.

Lydsegmentrelaterte feil i testmaterialet

Det var også tilfeller hvor det syntes å være åpenbare svakheter i segmenteringa. Med dette menes at lyder enten er for lite eller for mye trimmet, og liknende. Som tidligere nevnt utvikles en skjøtesyntese med enhetsutvelgelse ved å lese inn et manuskript og lagres i en taledatabase, jf. Kapittel 3.3., s. 10-15. Dette manuskriptet består hovedsakelig av setninger og fraser. Lydene en henter fra denne taledatabasen når syntetisk tale skal genereres, bør ha tydelig fonemisk annotering. Det vil si tydelige angitte start- og sluttpunkter som viser når en lyd starter og når den slutter i en streng med lyd. Men under den automatiske segmenteringa skjer det nokså ofte fonemisk feilannotering. Dette innebærer at et lydsegment kan bære en mindre eller større del av "nabofonemet" (eller ei "luftpause" dersom det aktuelle fonemet opptreer først eller til slutt i en lydstreng). Det ble observert enkelte tilfeller hos Kari som kunne synes å være et resultat av slik fonemisk feilannotering.

I setninga *En gift viste en lek som begrep* kunne det virke som om Kari uttalte *En gift vis(s)te en lek som begrep*. Her kunne det virke som om /i:/-lydsegmentet var kuttet i en så stor grad slik at det hørtes ut som en kort /I/-lyd. Til denne setninga var det ingen av de seks informantene som klarte å gjette seg fram til riktig ord. Siden både *å vise* (i betydninga *å la noen se; peke ut*) og *å vite* (i betydninga *ha kjennskap til; ha greie på*) er verb, som begge to vil gi den aktuelle setninga fullstendig syntetisk mening, vil det nok være mindre sannsynlighet for at informantene vil forsøke å tolke seg fram til ei anna betydning. Sammenliknet med Brage gjenga fem av seks informanter hele setninga korrekt, og med Stine kun én av seks informanter. Men som jeg skal komme nærmere tilbake til, ble det riktignok gjort funn som kunne indikere at informantene hadde et behov for å "dikte inn" ei mening dersom store deler av ei setning var uforståelig, jf. Kapittel 5.1.2, s. 111-112.

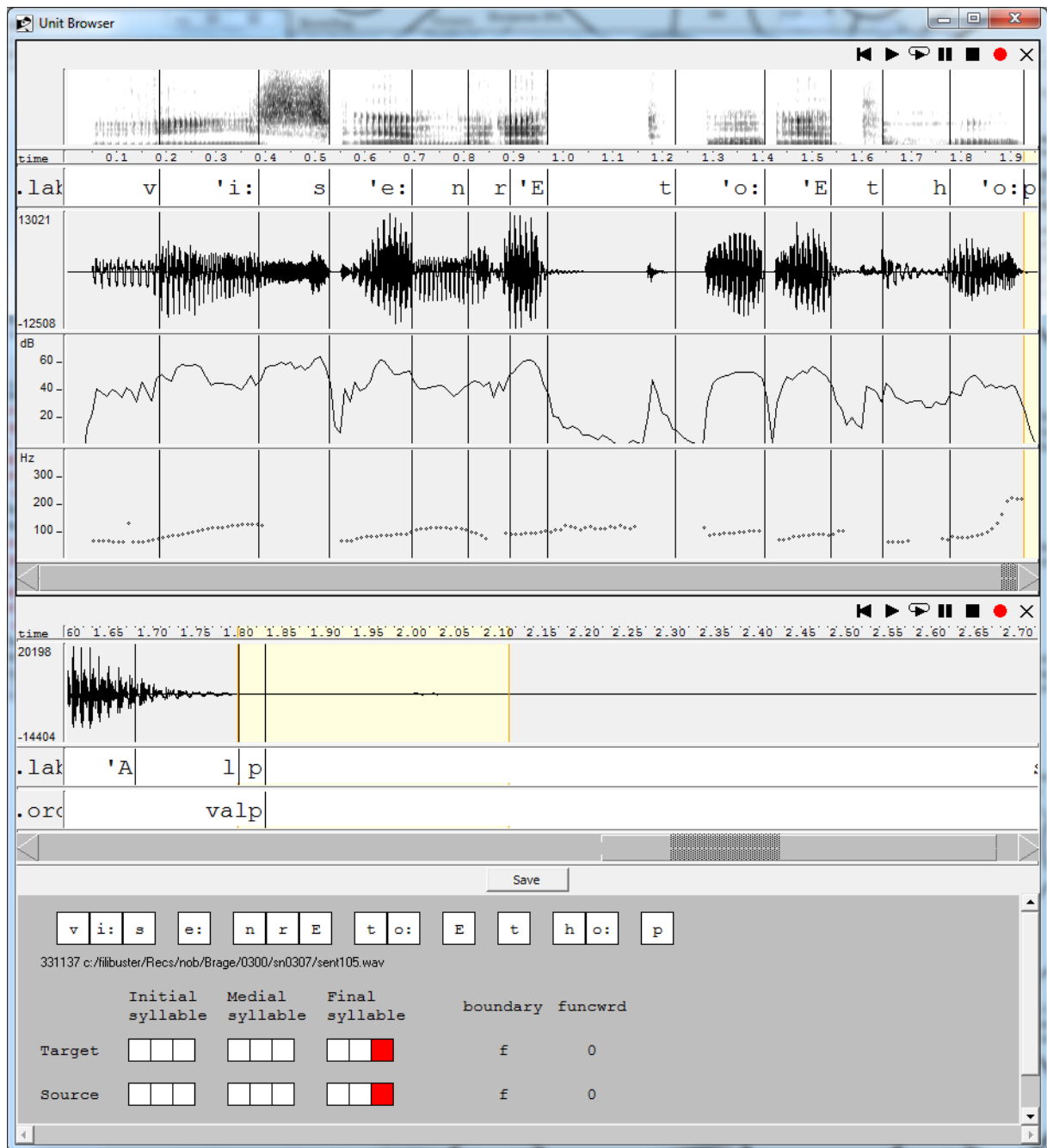
Et anna eksempel på noe som antakeligvis også var en segmenteringsfeil hos Kari, forekom i setninga *En glad kveld ga en start*. Kari uttalte denne setninga *En glad(e) kveld ga en start*. Dette var en grov feil. Her kunne det virke som om det hadde sneket seg inn en /@/-

lyd, dvs. en schwa-lyd, enten i /A:/-lydsegmentet eller i det foran- eller bakenforliggende lydsegmentet. En schwa er en urundet, midtre og sentral vokallyd, der midtre utgjør både trang-midtre og åpen-midtre. Schwa er den tyske skrivemåten til et opprinnelig hebraisk vokaltegn, som i IPA transkriberes /ə/. Eksempler på schwa kan være siste lyd i kodaposisjonen i ordet *å klare* (norsk), i kjerneposisjonen i ordet *again* ("igjen", engelsk), og i kjerneposisjonen i ordet *jeudi* ("torsdag", fransk). Sistnevnte eksempel er også et godt eksempel på at schwa-tegnet dekker vokaler over et temmelig stort område (Endresen, Simonsen og Sveen, 1996, s. 282-283). Tre av seks informanter gjettet seg fram til riktig uttale med Kari til tross for at det hadde sneket seg inn en schwa i et av lydsegmentene. Sammenliknet med Stine, gjettet fem av seks informanter seg fram til korrekt uttale, og med Brage gjettet samtlige seg fram til korrekt uttale.

Segmentrelaterte feil kunne også registreres hos Stine, men riktignok bare ett tilfelle. I setninga *En sint jobb skapte et liv* ble verbet *skapte* uttalt med en lang /A:/-lyd. Denne feilen kunne også ha vært forårsaket av tekstforprosessoren, som kunne ha oppfattet verbet som det refleksive verbet *å skape seg*. Men siden jeg heller ikke har tilgang til systemet som Stine er en del av, kan dette verken bekrefte eller avkrefte. Dersom man skulle anta at feilen var relatert til tekstforprosessoren, kunne en slik feil ha vært rettet opp ved å skille mellom verbene *å skape* (i betydninga *forme; danne*) og *å skape seg* (i betydning *gjøre seg til*). Dette kunne for eksempel ha blitt gjort ved å lage en regel hvor tekstforprosessoren søkte gjennom setninga etter det refleksive pronomenet *seg*. Funksjonaliteten i denne regel måtte være utformet slik at dersom det ikke forekom noe refleksivt pronomen, skulle første tilfellet brukes, og dersom tekstforprosessoren oppdaget det refleksivt pronomen (helst etterfulgt av verbet), ble dette verbet annotert som det refleksive verbet *å skape seg*. Men dette var en liten feil hos Stine, som kun én av seks informanter gjenga ukorrekt. Sammenliknet med Kari gjenga fem av seks hele setninga korrekt, og med Brage gjenga alle hele setninga korrekt.

Det ble kun observert ett tilfelle med en segmentrelatert feil hos Brage. Til gjengjeld var dette en vesentlig grov feil. I setninga *Vis en rett og et håp* var /p/-lydsegmentet kuttet helt vekk. Siden jeg har tilgang til Filibuster, kan jeg også spore opp hva denne feilen kommer av. Figur 13 nedenfor viser et enhetsleservindu i filibustersystemet. Dette vinduet gir fullstendig oversikt over all utdata til den aktuelle setninga, inkludert hvilke lydenheter som er konkatenerert og hvor /p/-lydsegmentet er hentet fra.

Figur 13. Bildet nedenfor viser et enhetsleservindu i filibustersystemet hvor utdata til setninga *Vis en rett og et håp* er presentert visuelt. Ovenfra og ned: første del framstiller utdatainformasjon for setninga ved hjelp av spektrogram, tidsakse, transkripsjon, pulsform, og kurve for styrke og toneleie. Andre del viser hvilken talesekvens /p/-lydsegmentet er hentet fra. Nederst vises enhetsutvelgelse.



I dette enhetsleservinduet er /p/-fonemet merket med gul utheving både i den øverste delen som angir utdatainformasjon til setninga *Vis en rett og et håp*. Som pulsformen angir er /p/-lyden helt vekk. Den fonetiske transkripsjonen /"h o: p/ (håp) består av en /p/ som er lenket til et lydsegment med stillhet. Den midtre delen av enhetsleservinduet viser hvilken

talesekvens i taledatabasen /p/-segmentet er lenket til. Figur 13 viser ikke hele talesekvens, som var *Hund, valp*. Det var /p/-fonemet i ordet *valp* som ble plukket ut og konkatenerert til *håp* i den aktuelle setninga. Som vist i den midtre delen av vinduet, ble dette fonemet for det første kuttet feil. Det er ikke så lett å se i denne figuren, men mot slutten i den gule hevinga kan det skimtes et lite utslag i pulsforma. For det andre var utslaget veldig svakt, dvs. at /p/-fonemet i *valp* ble uttalt så svakt av taledatabaseinnleseren, lyktes ikke autosegmenteringa å fange opp dette fonemet. I dette tilfellet ville det beste ha vært å fjerne hele talesekvensen fra taledatabasen, i stedet for å flytte sluttmarkøren som angir sluttposisjonen til lydsegmentet. Til tross for denne grove feilen klarte tre av seks informanter å gjengi hele setninga med korrekt uttale med Brage. Sammenliknet med Kari, gjenga kun to av seks informanter hele setninga korrekt, og med Stine kun én av seks informanter.

Lesehastighet og ordgrense

Som det kom fram av resultatene fra brukerundersøkelsen opplevde studenter som ikke var vant til talesyntese fra tidligere at det nokså sakte lesetempoet til Brage gjorde det lettere å få med seg meningsinnholdet i en tekst, mens de studentene som hadde større erfaring med å lese med talesyntese og foretrakk å lese hurtigere, ble frustrerte og irriterte over at Brage leste så tregt. Det er gjort studier som bekrefter at det synes å være en positiv korrelasjon mellom talekvalitet og lesehastighet. Blant anna i Sverige ble det gjennomført en studie i 1992 av Carlson m.fl. (1992, s. 63-66) og en annen studie i 1993 av Neovius og Raghavendra (1993, s. 21-30) som bekreftet denne korrelasjonen. Med bakgrunn i disse to studiene synes også mine funn å bekrefte at ei av årsakene til at Brage skåret høyere enn både Kari og Stine på en tydelighetstest, var hans moderate lesehastighet, jf. Kapittel 5.1.1, s. 59-63. Et moment som kan nevnes her er muligheten for at Kari og Stine ville ha skåret høyere i testen dersom utvalget mitt besto av informanter som hadde bedre erfaring med bruk av talesyntese.

Videre vil det være naturlig å diskutere lesehastighet og talekvalitet ved å studere eventuelle svake ordgrenser. Man tenker gjerne at det vil bli mindre ”luft” mellom ordene i ei ytring, desto hurtigere man snakker. Som tidligere illustrert leser Stine vesentlig hurtigere enn både Kari og Brage, jf. Kapittel 5.1.1, s. 59-63. I for eksempel setninga *Tvil en stein og en makt* syntes en informant at Stine leste *Tvilens stein og en makt*. Hun klarte altså ikke å definere de korrekte ordgrensene i denne setninga. Men det kunne ha vært interessant å ha

bedt samme informant om å gjengi samme setning som ble opplest med menneskelig tale. Gitt at en person leste opp denne setninga i samme tempo som Stine, ville denne informanten gjengitt setninga korrekt? Som man ser er det liten fonetisk forskjell i setningsdelene *Tvil en stein* og *Tvilens stein*. Pinker berører denne delen av talepersepsjon i boka *The Language Instinct* fra 1994 (Pinker, 1995, s. 158-163). Pinker hevder at vi hallusinerer ordgrenser. Ifølge han er all tale en illusjon. Med det mener han at vi hører tale som en streng med atskilte ord, hvor ingen av oss er i stand til å høre lyden av ei ordgrense. Vi hallusinerer ordgrensene ved å sammenlikne ei rad med lyder opp mot vår mentale ordliste. Med andre ord hevder Pinker at et ord løper inn i det neste som løper inn i det neste, osv., gjennom ei hel setning. Han beviser dette ved å vise at det ikke er noen ørsmå innslag med stillhet på samme måte som det er mellomrom mellom ord i en trykt tekst. *Tvil en stein* og *Tvilens stein* er ut ifra Pinker definisjon et eksempel på oronymi. Oronymer er strenger av ord som kan deles inn i ord på forskjellige måter. Her er noen engelske eksempler på oronymer som Pinker lister opp:

The good can decay many ways = The good candy came anyways.

The stuffy nose can lead to problems = The stuff he knows can lead to problems.

Slike oronymer vil bli vanskeligere å skille desto hurtigere en talesyntese leser. Brage har med bakgrunn i dette en fordel i og med at denne syntetiske stemmen leser sakte ved normalutgangspunkt, jf. Kapittel 5.1.1, s. 59-63. Det er mulig å heve og senke lesehastigheten med en DAISY-spiller, men selv om man kan senke lesehastigheten til Stine for eksempel ved slike oronyme tilfeller, vil de fremdeles være vanskelige å tyde. Grensa mellom ordene er den samme, selv om sekvensen spilles av i en lavere lesehastighet.

Det synes særlig å være setningene med imperativsstruktur (struktur nr. 3) som forårsaket oronyme feilgjengivelser. I tillegg til setninga *Tvil en stein* og *en makt* ble følgende funn gjort hos Stine. Samtlige tilfeller syntes å være oronyme i mindre eller større grad:

Varm et rom og en plan = Varmet rom og en plan.

Lik et tre og et kjøp = Like tre og et kjøp.

Drift en lov og et grep = Driften lå over et grep.

Sistnevnte tilfellet ble også gjengitt på tilsvarende ukorrekt måte av en informant som lyttet til samme imperative setning med Kari. I tillegg ble disse oronyme tilfeller med Kari observert:

Varmed et rom og en plan = Var med et rom og et plan.

Press et svar og en ferd = Presset svar og en ferd.

Ensett krav og en post = En sett krav og en post.

Kun én liknende feilgjengivelse ble observert med Brage:

Varmed et rom og en plan = Varmed rom og en plan.

Dette funnet synes å indikere en sammenheng mellom lesehastighet og vår evne til å angi ordgrensene på riktig sted.

Utydelighet forårsaket av prosodiske forhold

Neste punkt er prosodiske forhold som kan redusere tydeligheten til en talesyntese. Da taledatabasen som utgjør Brage i dag ble spilt inn, ble det sondret mellom vanlige setninger og spørresetninger. Med vanlige setninger menes fortellende setninger (Kulbrandstad, 1993, s. 197-199). Taledatabaseinnleseren leste inn de vanlige setningene ved at han konsekvent gikk ned i toneleie ved hvert punktum. Spørresetningene ble lest inn på en slik måte at han konsekvent gikk opp i toneleie ved slutten av hver setning. Hensikten var å tydeliggjøre forskjellen på spørresetninger kontra andre setninger overfor leseren. Man kunne også ha brukt prosodiske egenskaper til å sonde imperativssetninger fra andre setninger. For eksempel prosodiske trekk som ordtrykk og eventuelt også toneleie kunne ha blitt benyttet til dette formålet. Det ble likevel besluttet å lese inn imperativssetninger som vanlige punktum, men det finnes riktignok i dag talesynteser hvor det er eksperimentert med ekspressivitet (Loquendo, 2010).

Men selv om det er sondret mellom vanlige setninger og spørresetninger i taledatabasen, er denne toneleieforskjellen i dag fremdeles ikke implementert i Filibuster. Av den grunn behandler Filibuster i dag alle lydsegment i taledatabasen som om de alle er vanlige setninger. Dette betyr at når systemet skal generere uttale til spørresetninger og vanlige setninger, vil toneleiet ved utgangen til hver enkelt setning være vilkårlig. Dette er ei uheldig side ved Brage som særlig den tospråklige informanten fikk oppleve. De tolv spørresetningene ble for den tospråklige informanten fordelt på en slik måte at Brage leste seks setninger, mens Stine og Kari leste tre setninger hver. Det var kun ei av alle de tolv

spørresetningene hun klarte å forstå som en spørresetning. Det er mulig at resultatet ville ha vært annerledes dersom SUS-generatoren konstruerte spørresetninger med det andre spørreordet som var inkludert i testen (*hvorfor*), men det ble altså kun konstruert spørresetninger med spørreordet *når*. Det kunne ha vært spennende å følge opp dette med en ny test med flere tospråklige informanter, slik at man kunne bekrefte eller avkrefte om det er en relasjon mellom prosodiske trekk og talepersepsjon hos to- eller flerspråklige personer. Til resten av spørresetningene gjenga den tospråklige informanten *Når* som *Nå*. Ut ifra dette var informanten heller ikke i stand til å skille mellom vanlige punktum og spørresetninger verken ved uttalen til Brage, Stine eller Kari. Blant informantene med norsk som morsmål syntes det ikke å være noen utfordring å gjengi spørresetningene som spørresetninger. Dersom besvarelsen til den tospråklige informanten utelukkes og kun ses på besvarelsene til informantene med norsk som morsmål, ble fem av spørresetningene med Stine, tre av spørresetningene med Kari, og ingen av spørresetningene med Brage, gjengitt feil. Med bakgrunn i dette funnet kan en i videreutviklinga nedprioritere ei prosodisk korrigering av lydsegmentene for å skille mellom vanlige punktum og spørresetninger.

Et anna prosodisk trekk som kan være interessant å gå nærmere inn på er trykk og tonelag. Som tidligere nevnt har alle ord i norsk bokmål minst en trykksterk stavelse, og eventuelt én eller flere trykksvake (Kulbrandstad, 1993, s. 40-41). Dette markeres i den fonetiske transkripsjonen, som er lagret i uttaleleksikonet til Filibuster. Det samme gjelder for forskjellen i tonegangen, altså det som enten betegnes som tonelag 1 og tonelag 2, eller tonem 1 og tonem 2. Men selv om ordene har korrekt fonetisk informasjon i uttaleleksikonet, kan et system fremdeles generere uttale med manglende eller svakt trykk og ukorrekt tonelag, dersom det enten blir lest inn galt under innspillinga av taledatabasen, eller dersom talesekvensene i talebasen blir feilsegmentert på en eller annen måte.

Som nevnt kunne det virke som om enkelte hadde vansker med å definere ordgrensene i setningene de lyttet til, jf. Kapittel 5.1.2, s. 96-98. Kan ei av årsakene til dette være svakt trykk i ett eller flere av ordene, som inngikk i setningsdelen der ordgrensa var utydelig? Det ble registrert flere tilfeller der informantene trakk sammen ord til ett sammensatt ord. Til setninga *En glad kveld ga en start*, som ble lest av Stine, gjenga en informant *En gladkveld ga en start*. Siden jeg som nevnt ikke har tilgang til systemet som Stine er en del av, har jeg ikke mulighet til å finne noen eventuell feil. I denne setninga leste Stine ordene *glad* og *kveld* i samme lyse toneleiet uten noe særlig trykk på adjektivet *glad*. Det kan dermed tenkes at

taledatabasen til Stine ble lest inn med variasjon i setningsmelodi i mindre eller større grad. En slik variasjon kan ha både fordeler og ulemper for sluttbrukeren. En naturlig setningsmelodi med dynamikk og variasjon kan kanskje føre til mer naturlig uttale. Men man vil kunne trenge strengere regler i systemet for hvordan lydsegmentene skal velges ut. Slik som i dette eksemplet kan det kanskje blitt trukket ut lydsegmenter (til adjektivet *glad*) som ikke var egnet til neste ord (substantivet *kveld*) særlig med tanke på trykk og toneleie. For å oppnå en mest mulig naturlig syntetisk stemme, bør en under innspillinga av taledatabasen forsøke å lese inn alle setninger med den eksakt samme setningsmelodien. Man kan også gjøre dette med grunnlag i flere mønstre, ved for eksempel å lese inn setninger med samme syntaktiske struktur med den eksakt samme setningsmelodien. Å variere setningsmelodien til ulike syntaktiske strukturer representerer en risiko for en mindre naturlig talesyntese. For å være på den sikre sida og unngå uheldig enhetsutvelgelse, ble det som nevnt besluttet å lese inn samtlige setninger i det samme monotone toneleiet med Brage (bortsett fra å senke toneleiet ved utgangen av vanlige setninger og heve det ved utgangen av spørresetninger), jf. Kapittel 5.1.1, s. 56-59.

Dersom en ser bort ifra mindre sammenslåinger som *Press et = Presset* og *Tvil en stein = Tvilens stein*, og bare inkluderer tilfeller hvor to ord (hovedsakelig etter mønstret: adjektiv/substantiv + substantiv) ble slått sammen til ett meningsbærende sammensatt ord (hovedsakelig et substantiv), ble det registrert gjentatte tilfeller med feilaktig sammenslåing av ord. Dette ble observert hele 14 ganger hos Stine, tre ganger hos Kari, og ingen ganger i det hele tatt hos Brage. Ei sammenslåing som ble foreslått av fire informanter var *trykkfeil*. Den egentlige uttalen var *trygg feil*. Denne sammenslåinga ble også foreslått av én informant som lyttet til setninga med Kari. Et anna tilfelle av sammenslåing som ble repetert, var *sykunst*, hvor det egentlig ble uttalt *syk kunst*. Siden både kodaen til adjektivet *syk* og opptakta til substantivet *kunst* består av en "k", kan det være vanskelig for lytteren å gjengi korrekt ordgrense. Etter å ha lyttet nøye til disse setningene med Stine i ettertid, kunne det ikke sies å være noen svak eller gal betoning i testmaterialet. Dette funnet indikerer at det syntes å være en sammenheng med hurtigere lesehastighet og svakere avgjørelse av ordgrenser, jf. Kapittel 5.1.2, s. 96-98.

Siden sammenslåingene kun forekom i tilfeller med to enstavelsesord, som alene bare kan ha tonelag 1, kunne det ikke trekkes noen slutninger ut ifra testmaterialet til SUS-testen om at dette ble forårsaket av galt tonelag.

Når man skal måle talekvalitet på setningsnivå kan kanskje ”småord”, som for eksempel artikler og konjunksjoner, i første øyekast synes å være av mindre betydning for den helhetlige forståelsen av innholdet i ei setning. Funnene i SUS-testen synes å indikere hvor viktig disse er for den helhetlige forståelsen. Under innspillinga av taledatabasen til Brage ble det som nevnt besluttet at alle setningene skulle leses i samme lesehastighet, setningsmelodi og rytme, i den grad dette var mulig, jf. Kapittel 5.1.1, s. 50-51. Dette innebar at alle ordene i hver eneste talesekvens, som i dag utgjør stemmen Brage, ble lest med samme ”overtydelige” prosodi. Sammenliknet med menneskelig opplesing ville det være naturlig å lese mindre ord som determinativer, konjunksjoner og fyllord i høyere hastighet, og heller gå ned i lesehastighet og legge større ordtrykk på subjekter, verbaler og objekter i setningene. Slik prosodisk variasjon er noen av de mest effektive virkemidler en innleser har når han skal formidle en tekst.

Men som sagt syntes funnene i SUS-testen å indikere at ”småord” spilte ei vesentlig rolle for den helhetlige talekvaliteten til en talesyntese. I tabell 23 nedenfor er skårene fordelt mellom de tre talesyntesene ut ifra hvor mange determinativer (*en, et*) og konjunksjoner (*og, eller*) som ble gjengitt korrekt. Formålet med dette var å kunne si noe mer om betydninga av å lese inn taledatabasen til Brage med samme rytme og vekt på småord så vel som lengre ord.

Tabell 23. Antall ukorrekt gjengitte determinativer og konjunksjoner fordelt mellom de tre talesyntesene.

Talesyntese	Antall feil	Antall feil %
Brage	23	2,90 %
Kari	66	8,33 %
Stine	105	13,26 %

Det var 132 determinativer og konjunksjoner fordelt på de 60 setningene og seks gjennomføringer per talesyntese. Totalt utgjorde dette 792 determinativer og konjunksjoner. Det er nesten fem ganger så stor forskjell mellom Brage og Stine. Dersom jeg hadde valgt en test med meningsbærende setninger ville nok resultatet kanskje ha vært annerledes. Men dette er en av styrkene til SUS-testen: å kunne måle tydeligheten av determinativer og konjunksjoner. Særlig viktig er denne målinga med tanke på flerspråklige studenter. Det ble observert at enkelte informanter forvekslet mellom artiklene *en* og *et*. Denne forvekslinga

kunne i visse tilfeller føre til at ei setning fikk endret meningsinnhold, som for eksempel *en/et plan* (en: *forslag*; *utkast*; *grunnriss*. et: *nivå*), *en/et press* (en: *presset fold eller brett i buksene*. et: *trykk*), og *en/et rom* (en: *mannlig sigøyner*; *brennevin*. et: *værelse*; *volum*; *universet*). Dette funnet må forstås slik at ”småord”, særlig artikler, vil kunne være betydning for den helhetlige talekvaliteten. Dette kan ses i videre sammenheng med talepersepsjon og bruk av talesyntese blant syns- og lesehemmete som ikke har norsk som morsmål. Et særlig tilfelle som gikk igjen var sammenslåing av artikler og verb i imperativ. Dette medførte at hele den syntaktiske konstruksjonen ble endret, og dermed ble også meningsinnholdet i setninga vesentlig endret. Dette gjaldt for flere av informantene. Det er allerede nevnt at *Tvil en stein* ble gjengitt som *Tvilens stein*. Flere liknende funn ble gjort. Dette syntes å bekrefte viktigheten av ”småord” for den helhetlige tydeligheten, jf. Kapittel 5.1.2, s. 96-98.

Press et svar = Presset svar. Vis en rett = Visumrett. Varm et rom = Varmet rom.

Det som må tas med i denne beregninga er at dette ble målt med semantisk uforutsigbare setninger. Dersom setningene var meningsbærende ville ”småord” kanskje ha spilt ei mindre rolle. Men dette vil avhenge av at de kontekstuelle ordene blir uttalt tydelig.

Og med tanke på konjunksjonene var det ingen tilfeller hvor informantene hadde slått sammen disse med et ”nabo-ord”, slik som tilfellet var med artiklene. I stedet virket det som om det ble diktet inn andre ord dersom en konjunksjon var utydelig:

Spør et vann og en svikt = Spør et vann om en svikt (Stine).

Drift en lov og et grep = Driften lå over et grep (Kari, Stine).

Spør et vann og en svikt = Spør et vann orms svikt (Kari).

Det ble ikke observert ett eneste tilfelle der informantene feiltolket konjunksjonen *og* (det ble ikke generert ei eneste setning med konjunksjonen *eller* av SUS-generatoren) hos Brage, som hos Kari og Stine. Dette funnet vil indikere at lesehastighet og rytme synes å spille ei avgjørende rolle for forståelsen av konjunksjoner på setningsnivå.

Mindre alvorlige uttalefeil på setningsnivå

Med mindre alvorlige feil menes utydeligheter som synes å ha mindre påvirkning for talekvaliteten på setningsnivå. Dette vil først og fremst være tilfeller av feil samsvarsbøyning

av adjektiv, samt mindre alvorlige feilgjengivelser av preposisjoner. Særlig den tospråklige informanten gjenga adjektiv som for eksempel *frisk* (ubestemt, entall, maskulinum/femininum) i stedet for den korrekte formen *friskt* (ubestemt, entall, nøytrum). Og tilfeller med mindre alvorlige feilgjengivelser av preposisjoner var for eksempel at *mot* ble gjengitt som *imot*, og liknende. Ta for eksempel ei meningsbærende setning som *Jeg er imot monstermaster gjennom Hardangervidda*. Dersom denne gjengis *Jeg er mot monstermaster igjennom Hardangervidda*, må meningsinnholdet kunne sies å være tilnærmet det samme. Talekvaliteten ville i større grad ha blitt forringet dersom enten subjektet, verbalet eller et av subjektspredikativene var utydelig. Noen funn på slike mindre alvorlige feil der preposisjoner ble gjengitt ukorrekt i SUS-testen:

Et lån reiste fra et fritt skilt = Et land reiste ifra et frisk skilt (Stine).

En frykt fantes gjennom en bred sønn = En frykt fantes igjennom en bred sønn (Kari).

Et barn ba mot et trygt gram = Et barn ba imot et trygt lam (Brage).

At preposisjonsobjektet *lam* i eksemplet med Brage ble gjengitt galt, vil ha større betydning for talekvaliteten på setningsnivå enn at preposisjonen *mot* ble gjengitt som *imot*. Og at subjektet *lån* ble gjengitt som *land* i eksemplet med Stine ga setninga et helt annet innhold, sammenliknet med at preposisjonen *fra* ble gjengitt som *ifra*.

Vokaler versus konsonanter

SUS-testen gir også rom for å måle resultatene på fonemnivå. I dette kapitlet skal det forsøkes å kartlegge om visse lyder syntes å være vanskeligst å tyde enn andre. Var det enklere å tyde vokaler enn konsonanter? Hvilke type konsonanter var vanskeligst å tyde? Var det visse mønstre som gikk igjen? Og hva kunne dette være forårsaket av? I tabell 24 nedenfor er resultatene fordelt ved å regne ut hvor mange vokaler, diftonger og konsonanter som ble gjengitt galt hos de tre talesyntesene. Det er kun lydene som omfattes av et bokmålsnært norsk talemål, og ingen norske dialektale lyder, som vil bli diskutert. Konsonantene er listet opp etter de ulike artikulasjonsstedene. Med artikulasjonssted menes det stedet i talekanalen hvor det er størst innsnevring når man uttaler en lyd (Kulbrandstad, 1993, s. 23-36). Tabell 24 viser hvilke forskjellige artikulasjonssteder som er representert i bokmålsnært norsk talemål,

og antallet forekomster som de 60 setningene i SUS-testen utgjorde. For å forenkle framstillinga er dentaler, alveolarer og postalveolarer plassert i samme gruppe.

Tabell 24. Oversikt over hvilke norsk bokmålsnære konsonanter som er dekket av Filibuster med Brage, og antallet lyder som utgjorde datamaterialet til SUS-testen.

Artikulasjonssted	Beskrivelse innsnevring	Lyder	Antall
Leppelyder (bilabialer)	Mellom leppene	p, b, m	84
Leppe-tann-lyder (labio-dentaler)	Mellom tennene og underleppa	f, v	66
Tann-/gomlyder (dentaler, alveolarer, postalveolarer)	Mellom tungespissen/tungebladet og baksida av tennene og/eller gommen	t, d, n, s, l, r, S, l=, n=, tS	550
Bakgomlyder (retrofleksjer)	Mellom tungespissen/tungebladet og bakre del av gommen	t', d', n', s', l', l'=, n' =	15
Hardganelyder (palataler)	Mellom tungeryggen og den harde ganen	C, j	11
Bløtganelyder (velarer)	Mellom tungeryggen og den bløte ganen	k, g, N	84
Strupehodelyder (glottaler)	Mellom stemmebåndene i strupehodet	h	12

I utregninga av antall konsonanter ble konsonantiske lyder talt på grunnlag av normeringa som Berulfsen anbefalte (1969). Etter denne ble /d/-lyden i kodaen i ord som *med* (preposisjon), *tid* (substantiv), etc., talt, men ikke i adjektiv som *død*, *bred*, *glad*, *rød*, etc., eller i substantiv som *blod*, *hånd*, *kveld*, *stund*, *tråd*, etc. /g/-lyden i konjunksjonen *og* ble ikke talt. Ut ifra denne beregninga utgjorde datamaterialet totalt 822 bokmålsnære konsonanter.

Vokalene er delt inn etter deres tungeposisjon. Også diftongene er delt inn etter tungeposisjon, selv om diftongene endrer posisjon når de artikuleres (Kristoffersen, Simonsen og Sveen, 2005, s. 187). Felles for dette posisjonsskiftet er at de norske diftongene glir fra en nokså åpen vokal mot en trangere lyd (Kulbrandstad, 1993, s. 27-31). Datamaterialet utgjorde 471 norsk bokmålsnære vokaler og 13 diftonger med en fremre tungeposisjon. Det vil si at det ikke forelå noen diftonger med verken sentral eller bakre tungeposisjon.

Tabell 25. Oversikt over hvilke norsk bokmålsnære vokaler og diftonger som datamaterialet utgjorde.

	Fremre	Sentrale	Bakre
Vokaler	266	125	80
Diftonger	13		

Blant diftongene var det kun de fremre lydene /{I/ og /9Y/ som var representert i datamaterialet. I utregninga ble det gitt poeng for hver lyd som ble korrekt gjengitt i korrekt posisjon i forhold til foregående og etterfølgende lyder. For eksempel gjenga en informant verbet *tenkte* som *trengte*. Selv om gjengivelsen er gal på ordnivå, var samtlige lyder representert i korrekt posisjon i forhold til hverandre. Derfor ble det ikke trukket noen poeng i dette tilfellet. Det ble heller ikke hensyn til ordgrensene mellom ordene. For eksempel ble kun ett poeng trukket i dette tilfellet hvor en informant gjenga setninga *Et brev brukte en ro som økte* som *Et brev brukte en rose møkte* (/O/-lyden i *som* manglet). Hensikten med ei slik framstilling var å få en bedre oversikt over hvilke lyder som informantene hadde vanskeligheter med å fange opp.

For hver av talesyntesene var det mulig å skåre totalt (471 x 6 =) 2.826 poeng på vokalene, (13 x 6 =) 78 poeng på diftongene, og (822 x 6 =) 4.932 poeng på konsonantene. Tabell 26 nedenfor viser at datamaterialet utgjorde for få diftonger til å kunne si noe generelt om tydeligheten av disse.

Tabell 26. Resultatene fra SUS-testen fordelt mellom de tre talesyntesene på fonemnivå. Resultatene er framstilt i antall poeng og prosentvis i parentes.

	BRAGE		KARI		STINE	
ALLE VOKALER	2.805	99,26 %	2.701	95,58 %	2.529	89,49 %
Fremre	1.582	99,12 %	1.526	95,61 %	1.432	89,72 %
Sentrale	746	99,47 %	714	95,20 %	688	91,73 %
Bakre	477	99,38 %	461	96,04 %	409	85,21 %
ALLE DIFTONGER	77	98,72 %	77	98,72 %	63	80,77 %
Fremre	77	98,72 %	77	98,72 %	63	80,77 %

ALLE KONSONANTER	4.822	97,77 %	4.724	95,78 %	4.550	92,25 %
Bilabialer	496	98,41 %	490	97,22 %	451	89,48 %
Labio-dentaler	387	97,73 %	381	96,21 %	370	93,43 %
Dentaler, alveolarer, postalveolarer	3.221	97,61 %	3.154	95,58 %	3.062	92,79 %
Retrofleksjer	89	98,89 %	89	98,89 %	84	93,33 %
Palataler	65	98,49 %	64	96,97 %	61	92,42 %
Velarer	492	97,62 %	481	95,44 %	455	90,28 %
Glottaler	72	100,00 %	65	90,28 %	67	93,06 %

Informantene hadde større problemer med å gjengi alle lydene korrekt hos Stine, bortsett fra den glottale konsonanten /h/. I dette tilfellet skåret Stine to poeng mer enn Kari, mens alle informantene kunne gjengi samtlige glottale konsonanter hos Brage. Til setninga *En kreft reiste hos en høy sjel* kunne ingen av informantene tyde glottalen i artikulasjonen til Kari. Det var interessant at alle hørte en bilabial konsonant, /m/, selv om det ikke var noen bilabiale konsonanter verken blant de foregående eller blant de etterfølgende lydene. To av informantene som lyttet til denne setninga med Stine, hadde også problemer med å tyde glottalen /h/. Var det tilfeldig at én av disse to informantene også hørte en bilabial konsonant /m/ i stedet for glottalen?

Med hensyn til konsonantene skåret samtlige tre talesynteser svakere på de velare konsonantene. Det var to setninger som dro den samlede skåren til Brage ned. Til setninga *Et barn ba mot et trygt gram* ble den velare konsonanten /g/ i preposisjonsobjektet *gram* gjengitt tre ganger stumt og to ganger som en alveolar, /l/. Det ble også registrert ett tilfelle hvor velaren ble erstattet med en alveolar /l/ hos Kari og ett tilfelle hvor den ble erstattet med en labio-dental /f/ hos Stine. Etter å ha studert dette tilfellet i enhetsleservinduet til Filibuster, kunne dette problemet synes å være forårsaket av lydsegmentet som var lenket til alveolaren /r/. Dette lydsegmentet var kuttet med for seint startpunkt og inneholdt dessuten en del av en /A/-vokal. Og til setninga *Et rom vokste med en flott grad* ble velaren /g/ i preposisjonsobjektet gjengitt tre ganger stumt og én gang som en labio-dental, /f/ hos Brage. Dette ble også registrert med ett tilfelle hos Stine.

I tilfellene med velarene var det særlig noen som var interessante hos Kari og Stine. Til setninga *Når ville en lek glede en trygg feil*, som ble gjengitt feil én gang både hos Kari og

Brage, men hele seks ganger hos Stine. Her ble *trygg feil* gjengitt med ulike variasjoner av *trykkfeil*, *trykt feil*, etc. Dette ble tidligere vurdert som ei mulig ukorrekt betoning, jf. Kapittel 5.1.2, s. 100. Adjektivet *trygg* syntes å være vanskelig å tyde. Også til setninga *En trygg tekst endte et fag*, hvor adjektivet ble plassert ved sida av et anna substantiv, *tekst*, ble tilsvarende betoningsfeil registrert. Disse tilfellene kan ses i sammenheng med lesehastighet, jf. Kapittel 5.1.2, s. 59-63. Et anna tilfelle med tanke på velarer var at velaren /g/ i *gud* ble erstattet med en annen velar /k/ i *klut* i setninga *Tap en gud og en lek* hele tre ganger hos Stine. Til setninga *En gift viste en lek som begrep* ble velaren /k/ i *lek* feiltydet tre ganger hos Stine. Siden velaren var plassert i kodaposisjonen til ordet vil det være naturlig å anta at dette syntes å være forårsaket av svak eller utydelig innlesing under opprettinga av taledatabasen. Men dette kan ikke bekreftes ettersom jeg ikke kan etterprøve dette funnet med systemet som Stine er en del av.

Videre syntes det ikke å være noe repeterende mønster i de enkelttilfellene hvor palatale og retroflekse konsonanter ble gjengitt ukorrekt. I det følgende vil det bli redegjort for mønstrene som ble observert for gruppa med dentaler, alveolarer og postalveolarer. Ei rekke av feilgjengivelsene innenfor denne gruppa ble forårsaket av at alveolarene /n/ og /t/ i artiklene *en/et* ble forvekslet, jf. Kapittel 5.1.2, s. 102-103. Videre syntes alveolarene /t/ og /n/ igjen å være utfordrende å tyde, som i setninga *Tegn en kveld og en post*. Flere funn som indikerte at særlig alveolarene /t/, /n/ og /r/ syntes å være vanskelig å tyde, ble gjort. For eksempel i setninga *Et rom fantes mot et stort brev*, hvor alveolaren /t/ ble gjengitt som bilabialen /p/ tre ganger og én gang helt stum hos Brage. Denne feilen ble ikke registrert hos verken Kari eller Stine, sett bort ifra at tre informanter byttet ut retrofleksen /t/ i kodaposisjonen til adjektivet *stort* med en tynn alveolar /r/ hos Stine. Denne utydeligheten hos Brage var forårsaket av ei feiltranskribering i uttaleleksikonet til Filibuster. I uttaleleksikonet ble det funnet ei duplikatoppføring for nøytrumsadjektivet *stort* med kort /U/-vokal. Ettersom denne oppføringa hadde høyere rangering enn den normerte uttalen med lang /u:-vokal (Berulfsen, 1969, s. 307), ble dette plukket ut ved tekstnormaliseringa av Filibuster. Det var flere liknende tilfeller hos Brage hvor utydelige alveolarer ble oppfattet som bilabialer. I setninga *En natt spilte på en full grad* ble alveolaren /n/ i substantivet *natt* oppfattet som bilabialen /m/ hos to av informantene. Samme ord i samme setning ble oppfattet av en informant som *lapp* hos Kari. Og tilsvarende mønster ble funnet i setninga *En frykt fantes gjennom en bred sønn* hvor en informant gjenga preposisjonsobjektet *sønn* som *søm*.

Et anna repeterende mønster hos Brage var igjen forårsaket av alveolaren /r/. Fire informanter gjenga substantivet *vær* i setninga *Når måtte et vær plassere en fast stil?* enten som en alveolar /l/ eller som en retrofleks /n̠/. Igjen er dette et resultat av et utydelig /r/-lydsegment. Lyden syntes å være tilsynelatende distinkt uttalt i taledatabasen, men kan antakeligvis ha blitt for hardt trimmet under autosegmenteringa. Det ble også funnet gjentakende tilfeller hos Stine hvor informantene hadde vanskeligheter med å tyde alveolaren /r/. Substantivet *arv* i setninga *Når ville et skilt hente en redd arv?* ble gjengitt som *alv* av fire informanter. Alveolaren /r/ ble her altså byttet ut med en annen alveolar /l/.

Hos Kari var det særlig et tilfelle med alveolaren /n/ som skilte seg ut. Hele fem informanter gjenga denne alveolaren i substantivet *ånd* som en bilabial /m/ i setninga *Når ville en ånd leie en lik bønn?*. Til sammenlikning kunne ingen av informantene tyde hele eller store deler av samme setning med Stine. Det ble gjort flere funn der informantene forvekslet alveolaren /n/ og bilabialen /m/. Dette forekom også i setninga *En trygg tekst endte et fag*, hvor fire informanter byttet ut alveolaren /n/ med bilabialen /m/ i verbalet *endte*. Et liknende repeterende mønster ble observert hos Stine; i setninga *Press et svar og en ferd* alveolarene /r/ og /d/ byttet ut med bilabialen /m/ hele fem ganger. Og fire informanter hadde vanskeligheter med å gjengi alveolarene /d/, /l/, /n/, /s/ og /t/ i setninga *Del en stat og en tråd* hos Stine. Et slikt mønster med at utydelige alveolarer ble oppfattet som bilabialer syntes altså å være felles for de tre talesyntesene.

Det ble funnet enkelte tilfeller som kunne indikere at utydelige alveolarer også ble oppfattet som labio-dentaler. Særlig i setninga *Et valg bodde bak et trygt lys* ble adjektivet *trygt* oppfattet som *fryktet* hos Kari av tre informanter. Alveolaren /t/ ble her oppfattet som labio-dentalen /f/. Liknende eksempler ble funnet hos Kari ved av verbalet *tvil* ble oppfattet som *fil* i setninga *Tvil en stein og en makt*, og hos Stine i setninga *Når kunne et kurs selge en rød stein*, hvor alveolaren /s/ ble oppfattet som labio-dentalen /f/ av en informant.

Med tanke på utydelige labio-dentaler var det ingen repeterende mønstre hos verken Brage eller Kari. Og blant de enkelttilfellene som ble observert hos Stine, var det ingenting som kunne tyde på at utydelige labio-dentaler ble erstattet av andre konsonanter. For eksempel i setninga *Et rom fantes mot et stort brev* ble labio-dentalen /v/ i substantivet *brev* gjengitt av fem informanter som *bre* hos Stine.

Siste gruppe med konsonanter er bilabialer. Det var særlig et veldig stygt tilfelle hos Brage; i setninga *Vis en rett og et håp* hvor bilabialen /p/ var feilsegmentert. Hos Kari var det enkelte tilfeller som gikk igjen. I setninga *Et barn ba mot et trygt gram* gjenga fire informanter preposisjonen *mot* som preposisjonen *om*. Den bilabiale konsonanten ble gjengitt, men i feil posisjon i forhold til vokalen. Og i setninga *En redd mat ansatte en sum* falt bilabialen /m/ i substantivet *sum* bort for en informant og ble erstattet med alveolaren /n/ hos to informanter. Ellers syntes det ikke å være noe gjentakende mønster hos verken Brage eller Kari. Men hos Stine derimot syntes informantene å ha større utfordringer med å tyde bilabialene. Det ble registrert ei rekke tilfeller hvor nesten samtlige informanter enten manglet bilabialene eller erstattet dem med andre konsonanter. I de fleste tilfellene ble bilabialene hos Stine erstattet med alveolarer. For eksempel i setninga *Et barn ba mot et trygt gram* feilet fem informanter å registrere den andre bilabialen /b/ i verbalet *ba*. Ingen av informantene klarte å registrere dette verbalet. Kun én informant fanget opp begge bilabialene /b/, men gjenga lydsekvensen ukorrekt ved å flytte retrofleksen /nʰ/ til slutt slik *Et barbarbarn*. Et annet tilfelle hvor ingen av informantene klarte å tyde bilabialer var i setninga *En frisk sikt trengte en bønn* var alle informantene erstattet bilabialen /b/ i substantivet *bønn* med alveolaren /d/ hos Stine. Blant konsonantene var det bilabialene det syntes å være vanskeligst å tyde hos Stine.

Liknende funn ble registrert av Carlson, Granström og Nord under deres utvikling og evaluering av talesyntesystemet til Den kungliga Tekniska högskolan. De oppdaget at det var særlig bilabialen /m/, alveolaren /n/ og velaren /ŋg/ som syntes å være utfordrende å gjengi syntetisk (Carlson, Granström og Nord, 1990). Felles for disse konsonantene er at de alle tre er nasale. Det vil si at lydene uttales med åpen passasje ut gjennom nesa med deler av munnen fullstendig lukket (Endresen, Simonsen og Sveen, 1996, s. 208-221). Mine funn indikerer derimot at hos Brage syntes poengfordelinga å være temmelig jevnt fordelt blant de ulike lydene. Det ble observert størst variasjon hos Stine, jf. Tabell 26, s. 107. Hos Brage var det enkelte tilfeller hvor alveolaren /r/ ble feiltolket. En svært grov feil hos Brage var at bilabialen /p/ hadde falt helt bort. Men dette syntes å være enkelttilfeller.

Det ble ikke observert noen gjentakende mønstre på at informantene hadde utfordringer med å tyde velaren /ŋg/ slik som Carlson, Granström og Nord (1990) oppdaget. Dersom funnene rangeres etter artikulasjonsmåte, slik som Carlson, Granström og Nord, syntes det å være en viss forringelse av talepersepsjonen forårsaket av nasaler og plosiver hos alle tre talesyntesene. Plosiver, eller lukkelyder, er konsonanter som uttales ved at det dannes

et fullstendig lukke i deler av munnen, samtidig som passasjen på til nesa er avstengt (Kulbrandstad, 1993, s. 31-36).

Det er vanskelig å forstå hvorfor enkelte konsonanter syntes å være vanskeligere å tyde enn andre. Det er blitt pekt på helt konkrete svakheter i den fonetiske transkripsjonen, i lydsegmentene, og i taledatabasen hos Brage, og generelle antakelser om lesehastighet, betoning og feil ordklassifisering hos Kari og Stine. Siden et ”tekst-til-tale”-system er et komplekst system hvor svakheter eller åpenbare feil forplanter seg videre i de ulike delene av systemet, vil det derfor ikke alltid være lett å anslå eksakt hva en svakhet eller en feil kan være forårsaket av.

Hva med vokalene? Hvor tydelige var vokalene sammenliknet med konsonantene? Som tabell 26 viser varierte resultatene mellom de tre talesyntesene. Hos Brage syntes informantene å ha større utfordringer med å tyde konsonantene (97,77 %) sammenliknet med både vokalene (99,26 %) og diftongene (98,72 %). Hos Kari var resultatene temmelig likt fordelt mellom konsonantene (95,78 %) og vokalene (95,58 %). Kari skåret like høyt på diftongene som Brage. Men med Stine virket det som om informantene hadde større vanskeligheter med å tyde diftonger (80,77 %) og vokaler (89,49 %) framfor konsonanter (92,25 %). Hos Brage og Kari var resultatene nokså likt fordelt mellom de fremre, sentrale og bakre vokalene, mellom 99,12 og 99,47 % hos Brage og mellom 95,20 og 96,04 % hos Kari. Hos Stine syntes de bakre vokalene (85,21 %) å være vanskeligere å tyde enn de fremre (89,72 %) og de sentrale (91,73 %) vokalene.

Som tidligere nevnt var det særlig i de imperative strukturene at verbalet sammen med en nøytrumsartikkel forårsaket mange feil blant vokalene. Den fremre vokalen i artiklene *en* og *et* ble temmelig ofte gjengitt som en sentral vokal (schwa) ved at artiklene ble slått sammen med det imperative verbet, som for eksempel *Tvil en stein* = *Tvilens stein*. *Press et svar* = *Presset svar*. *Drift en lov* = *Driften lå*. *Vis en rett* = *Visen rettet*, m.fl. Denne feilgjengivelsen var representert i større eller mindre grad hos alle de tre talesyntesene. En annen feilgjengivelse som gikk igjen hos både Brage, Kari og Stine, var at den sentrale vokalen, schwa, uteble nokså ofte i spørresetningene som inneholdt verbalet *ville*. Dette verbalet ble tolket som *vil* gjentatte ganger. Spesielt for Stine gjenga fem av seks informanter substantivet *frykt* som *frukt* i setninga *En frykt fantes gjennom en bred sønn*. Ei anna vokalveksling som gikk igjen var /I/ og /i:/ → /Y/ og /y:/. Tre informanter gjenga substantivet *sikt* som adjektivet *syk(t)* med Stine i setninga *En frisk sikt trengte en bønn*. To andre tilfeller

hos Stine hvor vokalen *i* ble gjengitt som *y* var i setninga *En sang viste et salg som fikk*, hvor verbalet *fikk* ble gjengitt som *fyk/føyk* av fem informanter, og i setninga *Et grep trengte et smil som viste*, hvor verbalet *viste* ble gjengitt av tre informanter som *lyste*. Et anna repeterende mønster hos Stine med tanke på veksling mellom vokalene var /9/ og /2:/ → /E/ og /e:/, som for eksempel i setninga *Når ville en ånd leie en lik bønn?* hvor fem informanter gjenga substantivet *bønn* som *venn*.

Siden informantene hadde såpass store utfordringer med å gjengi vokalene hos Stine sammenliknet med Brage og Kari vil det være nærliggende å tro at lesehastighet påvirket poengfordelinga i en betydelig grad. Og når det gjelder spørresetningene hvor schwaen i *ville* ofte ikke uteble i talepersepsjonen til informantene, kan muligens også prosodien ha påvirket i disse tilfellene. Som tidligere nevnt sondres det fremdeles ikke mellom utsagnssetninger, spørresetninger og imperativssetninger hos Brage, jf. Kapittel 5.1.2, s. 98.

En mulighet for at mening "diktes inn"?

Det som var interessant å registrere var at enkelte av informantene syntes å skape mening i de utydelige delene av ei setning ved å erstatte utydelige ord med ord som var mer eller mindre meningsbærende i forhold til deler av eller hele setninga. I for eksempel setninga *En rett ansatte en drift som tok* var det flere av informantene som erstattet det direkte objektet *en drift* med *en bedrift*. Dette ble registrert to ganger hos Stine, én gang hos Kari og hele fire ganger hos Brage. Her vil det være nærliggende å anta at særlig verbalet *ansatte* kan ha medvirket til ei slik feiltolking. Et anna tilfelle som forekom tre ganger hos Brage og én gang hos Kari var at verbalet *eide* ble byttet ut med et anna verbal *veide* i setninga *Et gull eide en hud som etablerte*. Også her vil det være nærliggende å tro at subjektet *gull* kan ha ført til at flere gjenga verbalet feil. I setninga *Et barn ba mot et trygt gram* ble preposisjonsobjektet *gram* gjengitt som *lam* i tilfeller med Brage og ett tilfelle med Kari. Dette forekom gjerne også sammen med at preposisjonen *mot* ble erstattet med en annen preposisjon *om*: *Et barn ba om et trygt lam*. Et siste eksempel jeg vil trekke fram er setninga *En redd mat ansatte en sum*, hvor det ble registrert ett tilfelle hos Stine med frasen *En rød tomat*. At en tomat er rød gir mer mening enn at en mat er redd.

En enkel evalueringsmetode for talekvalitet på setningsnivå, slik SUS-testen er, vil ikke kunne måle muligheten for at folk kanskje forsøker å ”dike inn” mening. Allikevel er funnene interessante, men for å kunne gi svar på dette må man foreta ei psykologisk måling. Engstrand etterlyste i 1993 psykologisk måling, riktignok av talepersepsjon og konsentrasjon, i forbindelse med evaluering av talekvalitet på setningsnivå (Engstrand, 1993, s.35-44). Så vidt meg bekjent er det fremdeles ikke foretatt noen evaluering av talekvaliteten til en syntetisk stemme på setningsnivå, hvor det i tillegg er foretatt psykologisk måling.

Dette viser en av fordelene med SUS-testen, nemlig at denne testen ikke gir informanten noen semantisk tilleggsinformasjon. Informanten tvinges til å gjengi det han faktisk hørte. Ta for eksempel setninga *Kapteinen landet en Boeing 737 på Gardermoen flyplass*. Skulle et av ordene i denne setninga være utydelig vil allikevel en informant ha rikelig med semantisk informasjon til å gjette seg fram til det rette ordet. Flere forskere bekrefter denne fordelene med SUS-testen (Itahashi, 2000, s. 4; Li, Sityaev og Hao, 2007, s. 1350).

6 Oppsummering og konklusjoner

Nedenfor er talekvaliteten til Brage oppsummert på grunnlag av funnene fra undersøkelsene. Dette er framstilt etter de sentrale evalueringskriteriene. Konklusjoner om talekvaliteten er gitt på bakgrunn av målinger etter disse kriteriene. Kriteriet om velegnethet er det overordnede evalueringskriteriet i denne studien (Mariniak, 1993; King, 2007).

6.1 Tydelighet

I SUS-testen skåret Brage gjennomsnittlig høyere enn Kari og Stine både på setnings-, ord-, og fonemnivå. Det ble gjort ei rekke funn i SUS-testen. For det første synes det ikke å være noen av de fem syntaktiske strukturene som var vanskeligere å tyde med Brage enn andre. Brage taklet også homografene godt, sammenliknet med Kari og Stine. Ingen funn relatert til gal betoning av homografer, ble registrert hos Brage. Ett alvorlig funn ble gjort som skyldtes et svakt fonem i taledatabasen. Antakeligvis inneholder taledatabasen flere slike enkeltfeil.

Videre synes det å være en sammenheng mellom lesehastighet og evna til å definere ordgrenser. Tydeligheten av artikler synes også å spille ei vesentlig rolle for forståelsen på setningsnivå. Det synes å være en sammenheng mellom tydeligheten av artikler og en jevn, monoton rytme. Analysen av datamaterialet på fonemnivå indikerer at det ikke synes å være noen konsonantlyder eller vokallyder som var vanskeligere å tyde enn andre hos Brage.

Funnene fra SUS-testen må forstås som et uttrykk for hvor tydelig Brage synes å være generelt sett. Brage skåret jevnt høyere enn både Stine og Kari. Lesehastighet og monoton rytme synes å spille ei rolle. Den brukeropplevde tydeligheten derimot, sier SUS-testen ingenting om. Denne må vurderes ut ifra datamaterialet fra brukerundersøkelsen.

Hele 81 % av respondentene i brukerundersøkelsen oppga at Brage generelt sett enten hadde ei tydelig eller nokså tydelig artikulering. Dette kan jamføres med resultatet fra SUS-testen. Den spesielle tydeligheten derimot synes å være avhengig av hvordan studentene opplevde Brages mestring av fagterminologi som var sentral innenfor hver enkelt studieretning. Dette er alvorlig dersom det produseres studie bøker med Brage innenfor studieretninger som Brage ikke mestrer. Det bør settes i gang tiltak for å kunne anslå slike

dekningsgrader til hver enkelt studieretning. Studieretninger som er representert med en lav dekningsgrad, bør produseres med menneskelig innlesing inntil en akseptabel dekningsgrad kan dokumenteres.

I forhold til nynorsk, oppga 75 % at pensumlitteraturen deres ikke innholdt innslag av nynorsk. Uttaleleksikonet har i dag en svært lav dekningsgrad av nynorske ord (0,2 %). Men ettersom 21 % oppga at Brage tolket og behandlet innslag av nynorsk enten på en ok eller nokså god måte, vil ikke dette prioriteres i det videre utviklingsarbeidet. Og med tanke på behandling av engelske innslag, oppga omtrent halvparten (54 %) at Brage hadde en god engelskuttale. Dette må være et akseptabelt nivå, og vil heller ikke bli prioritert.

Det ble videre gjort funn i brukerundersøkelsen som indikerer en sammenheng mellom erfaring og talepersepsjon (Hjelmquist, 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Studenter med større erfaring med bruk av talesyntese over tid, vurderte den generelle tydeligheten av Brage bedre, sammenliknet med studenter med mindre erfaringer. Studentene antydte videre at Brage bør trenes mer på tolking og behandling av tall og talluttrykk, homografer og utenlandske egennavn.

6.2 Naturlighet

57 % av studentene likte selve stemmen til Brage godt eller nokså godt. Studenter som likte stemmen mindre godt eller ikke i det hele tatt, oppga også at de hadde liten erfaring med talesyntese. Selv om flere likte selve stemmen, oppga flere (45 %) at Brage var unaturlig eller nokså unaturlig. Ingen av disse hadde benyttet seg av talesyntese i mer enn seks år. På den andre sida oppga ingen av studentene som hadde brukt talesyntese i sju år eller lenger, at Brage var unaturlig.

At syntetisk tale liknet menneskelig tale mest mulig, var for de fleste (86 %) enten viktig eller nokså viktig. Enkelte (6 %) oppga riktignok at det ikke spilte noen rolle. Det var 13 % som oppga at de foretrakk både menneskelig og syntetisk tale, og 8 % som oppga at de foretrakk syntetisk tale. De som foretrakk syntetisk tale, begrunnet dette med muligheten for hurtigere tilrettelegging og større utvalg av studie bøker, samt at Brage ga dem elektronisk tilgang til tekstinholdet.

For nesten halvparten (42 %) spilte det ingen rolle hvilket kjønn en syntetisk stemme hadde, men for de øvrige syntes flere å foretrekke en mannlig stemme (42 %) framfor en kvinnelig (16 %). 59 % av respondentene oppga at prosodien til Brage enten var bra eller akseptabel, og 26 % ikke hadde noen særlige innvendinger. Dette må forstås slik at de fleste syntes at talekvaliteten til Brage var tilstrekkelig naturlig. Blant de som hadde ting å utsette på prosodien, påpekte 9 % at toneleiet var for mørkt, og 6 % at det var for liten variasjon i setningsmelodien. Leseflyten og rytmen til Brage ble vurdert av flere respondenter (25 %) som ujevn eller hakkete. Dette er ei generell utfordring for skjøtesynteser.

Opplevelsen av lesehastigheten til Brage var todelt. 49 % av studentene opplevde at Brage leste i et passelig tempo, mens 43 % syntes det enten ble for langsomt eller nokså langsomt. Studenter med større erfaring foretrakk en høyere lesehastighet, og oppga videre at de vanligvis pleide å skru hastigheten opp fra normalutgangspunktet under opplesing. Brage skiller seg ut i vesentlig grad fra de øvrige norske talesyntesene med hensyn til lesehastighet i normalutgangspunkt. Lesehastigheten til Brage lå 27,21 % over gjennomsnittet. Til tross for dette vil det ikke være aktuelt å justere normalutgangspunktet for lesehastigheten til Brage. Den moderate hastigheten imøtekommer de studentene som har behov for en mer saktelesende talesyntese. Lesehastigheten representerer det særlig karakteristiske for Brage, og utgjør dermed et alternativ til de øvrige norske talesynteser. Studenter som vil foretrekke en hurtigere talesyntese, kan velge blant de allerede eksisterende.

6.3 Behagelighet

34 % syntes det enten var behagelig eller nokså behagelig å lytte til Brage over tid, mens 30 % syntes det var ok, og 36 % enten ubehagelig eller nokså ubehagelig. Dette resultatet må forstås ut ifra flere hold. For det første syntes det å være stor variasjon innen individuelle preferanser, særlig i forhold til lesehastighet. Dette er observert i ei rekke evalueringsstudier (Furui, 2007, s. 22; Hjelmquist, Dahlstrand og Hedelin sitert i Cryer og Home, 2008, s. 8-9), men det synes ikke å være gjennomført noen studier som forsøker å finne årsaka til slik variasjon. Videre må dette funnet ses i sammenheng med hvor tydelig og naturlig hver enkelt opplevde at Brage var. I tillegg syntes vane å spille ei avgjørende rolle i forhold til hvor behagelig respondentene syntes talesyntese var generelt sett. Så mange som 17 av 24

respondenter (71 %), som ikke foretrakk å få studielitteratur tilrettelagt med syntetisk tale, hadde brukt en norsk talesyntese i mindre enn ett år. Dette indikerer at det er helt avgjørende å oppfordre til mer bruk av talesyntese for at studentene skal få mer positive brukeropplevelse med talesyntese (Hjelmquist, 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007).

Med hensyn til konsentrasjon oppga hele 83 % at de måtte anstrenge seg mer for å holde konsentrasjonen oppe ved syntetisk opplesing, sammenliknet med menneskelig. Studier viser at dette er et generelt problem for talesynteser (Goldstein, 1995; Engstrand, 1993; Axmear et al., 2005). Respondentene med mer erfaring syntes å anstrenge seg mindre enn respondentene med mindre erfaring.

6.4 Aksept

Om lag halvparten av studentene (45 %) foretrakk menneskelig tale framfor syntetisk. Det syntes å være en sammenheng mellom manglende eller liten erfaring med bruk av talesyntese og studentenes holdning til syntetisk opplesing generelt. Studenter som oppga at de hadde større erfaring med bruk av talesyntese, viste også en større aksept for tilbudet. Men riktignok oppga 45 % at de hadde stor tillit til Brage, 28 % at de hadde delvis tillit, mens 26 % oppga at de hadde liten eller ingen tillit. Ei manglende tillit ble begrunnet ut ifra både kriteriet om behagelighet, naturlighet og tydelighet. Særlig avgjørende for å vinne tilbake manglende tillit til Brage, bør det lages verktøy som kan bidra til å kartlegge og forbedre hvor velegnet denne talesyntesen er til å tolke og behandle sentral fagterminologi innenfor alle studieretninger.

6.5 Velegnethet

Hvor velegnet Brage er til å formidle studielitteratur, må ses i sammenheng med de øvrige evalueringskriteriene (Mariniak, 1993; Jekosch, 2005; King, 2007). Vurdert oppimot funnene fra brukerundersøkelsen og SUC-testen synes Brage å være generelt egnet til å tolke og behandle de tekstuelle utfordringer som særpreger norsk studielitteratur. 15 respondenter foretrakk Brage som oppleser. Dette var fem ganger så mange sammenliknet med de øvrige talesyntesene. Men hele ni av de 15 respondentene som foretrakk Brage, oppga at de hadde

benyttet en norsk talesyntese i mindre enn ett år. Fire av disse ni oppga at de hadde kjennskap til andre talesynteser. Dermed er det ikke nødvendigvis en aksept for Brage som de uttrykte, men en aksept for talesyntese generelt. Man kan heller ikke avskrive at de også likte Brage.

I brukerundersøkelsen svarte 47 % at de syntes det var lett eller nokså lett å få med seg meningsinnholdet i en tekst, som ble opplest av Brage, mens 30 % syntes det enten var vanskelig eller nokså vanskelig.

Kriteriet om velegnethet må ses i sammenheng med det tiltenkte bruksområdet til Filibuster. Denne talesyntesen skal brukes til å produsere stort sett all type studielitteratur. Studielitteratur spenner over et vidt område av fagfelt og funn fra brukerundersøkelsen antyder at Brage synes å være mindre velegnet til å produsere studie bøker innenfor visse studieretninger. Det må utvikles verktøy for å kunne kartlegge dekningsgraden av fagterminologi innenfor hver enkelt studieretning. Før en akseptabel dekningsgrad kan dokumenteres, anbefales menneskelig innlesing framfor produksjoner med talesyntese, jf. Kapittel 7, s. 118.

7 Veien videre

I det følgende er det foreslått enkelte tiltak til det videre arbeidet med Filibuster med Brage på grunnlag av funnene i denne studien.

Det ble nevnt i diskusjonen av resultatene fra brukerundersøkelsen at økonomiske og juridiske fagtekster kan synes å representere syntaktiske konstruksjoner, som skilte seg ut fra øvrig faglitteratur. Det vil ikke være tilstrekkelig med å tilføye sentral fagterminologi innenfor disse studieretningene i uttaleleksikonet. Regelsettet i tekstforprosessoren til Filibuster må dessuten trenes oppimot slike syntaktiske utfordringer.

Et annet moment som ble nevnt i denne studien var at SUC-tagget, som brukes til å annotere registreringer i uttaleleksikonet morfologisk, ikke sonderer mellom hjelpeverb fra andre verb. Ved å tilføye en ny tagg til det eksisterende tagget, vil man kunne lage noen generelle morfologiske analyseregler for hjelpeverb i tekstforprosessoren. I dag kan kun ortografisk analyseregler lages. Ved å bruke morfologisk informasjon vil man kunne fange opp generelle tilfeller, og dermed slippe å skrive nye regler for hvert eneste ortografiske tilfelle.

Ettersom 45 % av respondentene i brukerundersøkelsen syntes Brage var unaturlig eller nokså unaturlig, bør det også prioriteres tiltak som vil føre til at Brage blir mer naturlig. En mulighet er frasespleising, slik som IBM har gjort med sine talesynteser (Kvale, 2006, s.22). Ved hjelp av frekvensinformasjon kunne man ha kartlagt hvilke fraser og uttrykk som går igjen i det totale produksjonsgrunnlaget til Brage, og brukt denne informasjonen som manuskript til en supplerende taledatabase, der høyfrekvente fraser kan plukkes ut i sin helhet, i stedet for å konkatenerer på fonemnivå. Dette vil også kunne forbedre tydeligheten, ettersom naturlighet og tydelighet er nært knyttet til hverandre.

Det bør utvikles verktøy som gir oversikt over dekningsgrad av sentral fagterminologi i uttaleleksikonet til Filibuster. Dette vil i første fase innebære at NLB må tilføye tilleggsopplysninger til hver enkelt DAISY-bok for å kunne ta i bruk et slikt verktøy. Studiebøkene må inneholde søkbar informasjon om hvilke studieretninger de tilhører. Dette kan være et tid- og ressurskrevende arbeid, men på sikt vil dette kunne effektivisere det videre arbeidet med uttaleleksikonet i større grad enn tidligere praksis med tilfeldige lytteprøver.

Kanskje ei mellomløsning ville ha vært en bokstaveringsfunksjon i DAISY-spilleren? En slik bokstaveringsfunksjon kan løses ved at en innebygd talesyntese bokstaverer utydelige ord. En slik funksjon vil sikre tilgang til teksten. Det amerikanske selskapet IRTI (*Innovative Rehabilitation Technology Inc.*) har laget en slik bokstaveringsfunksjon i spilleren sin, eCclipseReader-EyeNet. Men den funksjonen kan bare brukes i rene tekstfiler og ikke i ei DAISY-fulltekstbøker. Så vidt meg bekjent mangler en slik funksjon i spilleren som NLB i dag anbefaler til lånerne, Dolphin EasyReader (versjon 3.01). I betingelsene som DAISY-konsortiet utarbeidet i 2008 for DAISY-spillerne var ett av kravene nettopp en slik bokstaveringsfunksjon (Kerscher, Kahl og Sutton, 2008). I punkt 2.3.3 heter det at en slik funksjon skal inkluderes i bøker som produseres med talesyntese. DAISY-standarder gir rom for dette, men ikke enkelte spillere. En slik funksjon ville også være behjelpelig for synshemmete for å sikre korrekt kildehenvisning ved oppgaveskriving. Ta for eksempel navnet Smith, som også kan skrives *Smitt*, *Schmitt*, eller *Schmidt*.

Funn fra brukerundersøkelsen og andre tidligere studier viser en sammenheng mellom talepersepsjon og erfaring (Hjelmquist, 1990; Vankatagiri, 1994; Francis, Nusbaum og Fenn, 2007). Dette er viktig kunnskap som bør brukes. Det kunne ha blitt opprettet et brukerorganisert forum der syns- og lesehemmete studenter sammen med andre lesehandikappede kunne ha delt av hverandres erfaringer. Kun gjennom erfaring vil man oppdage ny funksjonalitet, som vil gjøre tekst mer tilgjengelig. Et slikt forum kunne enten ha blitt organisert gjennom arrangementer i regi av NLB. Alternativt kunne man ha opprettet et interaktivt forum på Internett. Formålet med et slikt forum vil også være å oppnå større aksept blant studentene til bruk av talesyntese.

Viktig informasjon som kan effektivisere det videre utviklingsarbeidet med Filibuster med Brage, er frekvensinformasjon. Frekvensinformasjon kan brukes på flere måter. Før ei studiebok produseres kan det opprettes ei frekvensliste basert på tekstinnholdet i den. Ved å sammenlikne denne frekvenslista med uttaleleksikonet kan man effektivt kartlegge høyfrekvent fagterminologi. Med ei stram tidsramme vil det være begrenset hvor mange uregistrerte ord man rekker å føre opp i uttaleleksikonet før den produseres, men det kan i alle fall være en mulighet til å rette opp noen utydelige forekomster. Et slikt verktøy kan også brukes til å søke etter skrivefeil i tekstinnholdet, og eventuelt skannefeil ved den elektroniske konverteringa. Det kan gjøres ved å sammenlikne frekvenslista oppimot uttaleleksikonet.

Funn fra SUS-testen antydte at det syntes å være opptakta i de utydelige ordene som ofte ble ukorrekt gjengitt. Rimet til et ord syntes oftere å bli gjengitt korrekt. Kan disse funnene synes å være svake innleste fonem fra taledatabasen? Hvis så bør det lages verktøy for å søke etter slike svake innleste lyder. Eventuelt nye feil bør sammenliknes med eldre som grunnlag til å trene søkealgoritmen, slik at man kan få et bedre søkeverktøy som igjen kan avdekke flere feil. Dette kan gi funn som vil være av betydning for det videre arbeidet med korrigering av lydsegmentene i taledatabasen til Brage.

Videre vil det være ønskelig å gjennomføre nye liknende undersøkelser etter at systemet og stemmen er forbedret oppimot funn som ble gjort i denne studien. En ny undersøkelse med tilsvarende innhold vil kunne gi et sammenliknbart datamateriale. Dette vil også kunne gi muligheten til å måle effektiviteten av tiltakene som her foreslås. I tillegg bør det vurderes å samle inn mer kvalitative data enn hva som ble gjort gjennom denne undersøkelsen. Dette vil bidra til å kartlegge nye og ukjente sider med studenters opplevelser med Brage.

8 Svakheter ved studien

Enkelte spørsmål i spørreskjemaet kunne synes å være tvetydig eller upresist formulert. Det var også enkelte spørsmål der svaralternativene var formulert slik at de ikke fanget opp alle sider ved studiesituasjonen. De ukjente sidene som spørreskjemaet ikke tok hensyn til, kom først fram i dialog med hver enkelt student. Dette kunne kanskje ha vært unngått dersom det hadde vært gjennomført en pilotundersøkelse med noen få studenter. Spørreskjemaet manglet videre et spørsmål som spurte direkte om hva slags holdning studentenes hadde til talesyntese generelt. Dette burde ha vært inkludert som en sentral bakgrunnsvariabel.

De fleste besvarelsene ble samlet inn over telefon, slik at studenter som ikke kunne benytte seg av de forhåndsdefinerte svaralternativene, kunne avklare dette sammen med datainnsamlerne. Det var også tilføyd et eget notatfelt til hver del i spørreskjemaet, som studentene brukte til å komme med utfyllende kommentarer. Studentene som fylte ut skjemaet selv, ble også bedt om å bruke dette notatfeltet i forkant av datainnsamlinga.

Enkelte respondenter som fylte ut skjemaet selv, lot være å svare på spørsmål med åpne svar. Dette var en risiko å ta ved å la respondentene fylle ut spørreskjemaet på egenhånd. Det kunne også ha blitt benyttet spørsmål med forhåndsdefinerte svaralternativ. Men i visse spørsmål var det mindre heldig å bruke svaralternativ, som for eksempel opplevelsen av prosodien til Brage. Her kunne viktige sider ved prosodien utelates på grunn av manglende svaralternativ. Samtidig ville forhåndsdefinerte svaralternativ kunne legge føringer og hindre egen refleksjon hos studentene. Ideelt sett burde all datainnsamling ha blitt gjennomført over telefon, for å unngå at enkelte besvarelser måtte forkastes.

Med tanke på SUS-testen kunne andre talesynteser enn Kari og Stine synes å være like aktuelle som sammenlikningsgrunnlag. Det var ønskelig å velge to andre talesynteser som studentene syntes å foretrekke. Dette valget ble tatt ut ifra resultatet fra brukerundersøkelsen på bakgrunn av spørsmålet: ”Dersom du kunne velge en av disse norske talesyntesene som oppleser for din studielitteratur, hvilken ville du ha foretrukket?” Valget kunne like gjerne ha falt på Nora eller Vegard, ettersom tre respondenter foretrakk Kari og Stine, og to respondenter foretrakk Nora og Vegard.

Tabelliste

Tabell 1.

Oversikt over sentrale norske talesynteser. 9

Tabell 2.

De fem syntaktiske strukturene i SUS-testen med engelske og svenske eksempler. 23

Tabell 3.

Oversikt over kartlegginga av utvalget til brukerundersøkelsen. 31

Tabell 4.

Oversikt over hvor mange oppføringer som kreves innen hver ordklasse og syntaktiske kategori, og hvor mange ord som ble benyttet. 39

Tabell 5.

De fem syntaktiske strukturene med norske eksempler. 40

Tabell 6.

Fordeling av utvalget etter hvor naturlig respondentene syntes Brage var, sammenliknet med en menneskelig innleser, ut ifra hvor lenge respondentene hadde benyttet en norsk talesyntese. N=53. 44

Tabell 7.

Fordeling av utvalget etter hva respondentene syntes om å motta studielitteratur som var produsert med syntetisk tale, sammenliknet med hvilket lesehandikap respondentene hadde. Frekvens. N=53. 47

Tabell 8.

Krysstabell som viser resultatene fordelt på hvor godt respondentene liker selve stemmen til Brage og hvor lenge de har brukt norsk talesyntese. N=53. 50

Tabell 9.

Resultatene fordelt etter hvilke oppfatninger respondentene hadde om prosodien til Brage (rader), sammenliknet med hvor lenge de hadde benyttet seg av en norsk talesyntese (kolonner). Frekvens. N=53. 51

Tabell 10.

Oversikt over differansen i spilletid mellom Brage og andre norske kommersielle talesynteser. Målingene er gjort ut ifra talesyntesenes normalhastighet. 60

Tabell 11.

Fordeling av utvalget etter hva respondentene syntes om lesehastigheten til Brage i normalhastighet (kolonner), sammenliknet med hvordan de syntes det var å få med seg meningsinnholdet i en tekst, som ble opplest av Brage (rader). N=53. 61

Tabell 12.

Fordeling av utvalget etter hva respondentene syntes om lesehastigheten til Brage i normalhastighet (rader), sammenliknet med hvilken norsk talesyntese de foretrakk (kolonner). N=53. 62

Tabell 13.

Fordeling av utvalget etter hvor tydelig studentene syntes den generelle artikuleringa til Brage var, sammenliknet med hvilken norsk talesyntese de foretrakk som oppleser av studielitteratur. N=53. 64

Tabell 14.

Fordeling av utvalget etter hvor ofte respondentene syntes uttalemessige utydigheter forekom. Frekvens. N=53. 65

Tabell 15.

Oversikt over hvordan respondentene opplevde at Brage mestret faguttrykk innenfor deres studieretning, sammenliknet med hvilken studieretning respondentene deltok innenfor. 68

Tabell 16.

Fordeling av utvalget etter om respondentene opplevde at de måtte anstrenge seg mer for å holde konsentrasjonen oppe da de lyttet til DAISY-bøker, som ble lest opp av Brage, sammenliknet med DAISY-bøker med menneskelig opplesing. Resultatene er videre sammenliknet med hvor lenge studentene hadde benyttet seg av en norsk talesyntese. N=53. 75

Tabell 17.

Fordeling av utvalget etter hvordan respondentene syntes det var å få med seg meningsinnholdet i en tekst, som ble lest opp av Brage (rader), sammenliknet med hvordan de syntes Brage mestret faguttrykk innenfor studieretninga si (kolonner). N=53. 78

Tabell 18.

Krysstabell som viser hvor lenge studentene hadde benyttet seg av en norsk talesyntese og hvilken talesyntese de foretrakk. 79

Tabell 19.

Fordeling av utvalget etter hvilken studieretning respondentene studerte innenfor (rader), sammenliknet med hvilken grad av tillit de hadde til Brage som formidler av studielitteraturen deres (kolonner). N=53. 84

Tabell 20.

Resultatene fra SUS-testen fordelt som gjennomsnittligsskår per informant og totalskår for seks informanter. 87

Tabell 21.

Resultatene fra SUS-testen fordelt på hver av de fem syntaktiske strukturene. 88

Tabell 22.

De fem minst tydelige setningene hos hver av de tre talesyntesene. 91

Tabell 23.

Antall ukorrekt gjengitte determinativer og konjunksjoner fordelt mellom de tre talesyntesene..... 101

Tabell 24.

Oversikt over hvilke norsk bokmålsnære konsonanter som er dekket av Filibuster med Brage, og antallet lyder som utgjorde datamaterialet til SUS-testen. 104

Tabell 25.

Oversikt over hvilke norsk bokmålsnære vokaler og diftonger som datamaterialet utgjorde..... 105

Tabell 26.

Resultatene fra SUS-testen fordelt mellom de tre talesyntesene på fonemnivå. Resultatene er framstilt i antall poeng og prosentvis i parentes. 105

Figurliste

Figur 1.

Figuren viser veien fra tekst til tale. Etter modell opprinnelig laget av Tånnander og Sjölander ved TPB. 12

Figur 2.

Fordeling av utvalget etter hvor mange respondenter som hadde produksjonsrett og lånerett (til venstre), og etter lesehandikap (til høyre). Prosenter. N=53. 33

Figur 3.

Fordeling av utvalget etter hva respondentene foretrakk av syntetisk eller menneskelig tale ved opplesing av studielitteratur. Prosenter. N=53. 46

Figur 4.

Eksempel på stavellesstruktur..... 55

Figur 5.

Eksempel på hvordan Filibuster velger ut lydsegmenter til setninga *Ikke alle tjenester i arbeidslivet er motivert ut fra økonomisk egeninteresse*. Den første linja viser resultatet fra konkateneringa, mens den nederste linja viser hvilke innleste sekvenser hvert segment er hentet fra. 58

Figur 6.

Fordeling av utvalget etter hvordan studentene syntes Brage mestret innslag av nynorsk. Prosenter. N=53..... 71

Figur 7.

Fordeling av utvalget etter hvordan studentene syntes Brage mestret innslag av engelsk. Prosenter. N=53..... 73

Figur 8.

Fordeling av utvalget etter hvordan respondentene opplevde det var å lytte til studielitteratur, som ble opplest av Brage, over tid. Prosenter. N=53..... 74

Figur 9.

Fordeling av utvalget etter om studentene opplevde at de måtte konsentrere seg mer for å holde konsentrasjonen oppe da de lyttet til DAISY-bøker, som er produsert med Brage, sammenliknet med DAISY-bøker, som er produsert med menneskelig innlesing. Resultatene er videre sammenliknet med hvor lenge studentene hadde benyttet seg av en norsk talesyntese..... 76

Figur 10.	
Fordeling av utvalget etter hvordan respondentene syntes det var å få med seg meningsinnholdet i en tekst, som ble opplest med Brage. Prosenter. N=53.....	77
Figur 11.	
Fordeling av utvalget etter hvilket kjønn og stemmeleie som var å foretrekke til en syntetisk stemme. Frekvens. N=53.	82
Figur 12.	
Utdrag fra frekvensanalysen av stemmen til mannen bak den syntetiske stemmen Brage.....	83
Figur 13.	
Bildet nedenfor viser et enhetsleservindu i filibustersystemet hvor utdata til setninga <i>Vis en rett og et håp</i> er presentert visuelt. Ovenfra og ned: første del framstiller utdatainformasjon for setninga ved hjelp av spektrogram, tidsakse, transkripsjon, pulsform, og kurve for styrke og toneleie. Andre del viser hvilken talesekvens /p/-lydsegmentet er hentet fra. Nederst vises enhetsutvelgelse.	95

Litteraturliste

Programvare, lister og annet materiale som ble brukt i forbindelse med SUS-testen kan lastes ned fra <http://www.teksttiltale.no/Evalueringsmetoder/SUS/>.

Acapela Group 2009, (25.02.2009), *Infovox iVox 1.3: new voices available!*, Acapela Group, tilgjengelig fra: <<http://www.acapela-group.com/infovox-ivox-1-3--new-voices-available-2043-speech-synthesis.html>>, [19.05.2010].

Acapela Group 2004, (12.10.2004), *Hear Acapela's new high quality natural sounding Norwegian voice at ID Dagarna in Stockholm*, [online], Acapela Group, tilgjengelig fra: <<http://www.acapela-group.com/hear-acapela-s-new-high-quality-natural-sounding-norwegian-voice-at-id-dagarna-in-stockholm--10-speech-synthesis.html>>, [19.05.2010].

Ahrenberg, L., Fenstad, J. E., Kvale, K., Maegaard, B. og Mühlenbock, K. 2001, *KUSTI – Kunnskapsutvikling for norsk språkteknologi. Programplan*, Norges forskningsråd, tilgjengelig fra: <<http://www.forskningsradet.no/servlet/Satellite?blobcol=urldata&blobheader=application%2Fmsword&blobheadervalue1=Content-Disposition%3A%3B+filename%3Dkunstiplan.doc&blobkey=id&blobtable=MungoBlobs&blobwhere=1274460314443&ssbinary=true>>, [22.04.2010].

Allen, J., Hunnicutt, S. og Klatt, D. 1987, *From Text to Speech: The MITalk System*, Cambridge University Press, Inc., Cambridge.

Amdal, I., Strand, O. M., Almqvist, J. og Svendsen, T. 2008, *RUNDKAST: An Annotated Norwegian Broadcast News Speech Corpus*, Proceedings of LREC 2008, Marrakech, tilgjengelig fra: <http://www.iet.ntnu.no/~torbjorn/Papers/2008_LREC_Rundkast.pdf>, [24.04.2010].

Amdal, I., Johnsen, M. H. og Svendsen, T. 2006, *Log Likelihood Ratio Based Annotation Verification of a Norwegian Speech Synthesis Database*, Dpt. Of Electronics and Telecommunications, NTNU, tilgjengelig fra: <http://www.iet.ntnu.no/~torbjorn/Papers/2006_Norsig_LLRR.pdf>, [24.04.2010].

Amdal, I. og Svendsen, T. 2006, *FonDat1: A Speech Synthesis Corpus for Norwegian*, Proceedings of LREC-2006, Genoa, Italy, tilgjengelig fra: <http://www.iet.ntnu.no/~torbjorn/Papers/2006_LREC_fondat1.pdf>, [24.04.2010].

Amdal, I. 2005, *FonDat1: En database for utvikling av verktøy for norsk datadrevet skjøtesyntese*, MONS11, 11. møte om norsk språk, Universitetet i Bergen, s. 13-14,

- tilgjengelig fra: <<http://www.hf.uib.no/i/Nordisk/MONS11/Konferansehefte.pdf>>, [24.04.2010].
- Axmear, E., Reichle, J., Alamsaputra, M., Kohnert, K., Drager, K. og Sellnow, K. 2005, *Synthesized Speech Intelligibility in Sentences. A Comparison of Monolingual English-Speaking and Bilingual Children*, i *Language, Speech, and Hearing Services in Schools*, Vol. 36, s. 244-250, tilgjengelig fra: <<http://lshss.asha.org/cgi/content/abstract/36/3/244>>, [20.10.2010].
- Barraud, M. 2010, *CNIB – What We Believe – Nationwide Accessible Library Service to Print Disabled Persons*, [online], The Canadian National Institute for the Blind, tilgjengelig fra: <<http://www.cnib.ca/en/about/who/believe/Pages/print-disabled.aspx>>, [06.11.2010].
- Benoît, C., Grice, M. og Hazan, V. 1995, *The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences*, *Speech Communication*, Vol. 18, 1996, s. 381-392, tilgjengelig fra: <<http://www.phon.ucl.ac.uk/home/val/sus%20paper.pdf>>, [29.04.2010].
- Bentzen, A. (07.09.2010), *Skatteetaten gikk på ellevill snakkesmell – hør klippet*, [online], TV 2 Nyhetene, tilgjengelig fra: <<http://www.tv2nyhetene.no/innenriks/skatteetaten-gikk-paa-ellevill-snakkesmell-hoer-klippet-3283807.html>>, [08.09.2010].
- Berulfsen, B. 1969, *Norsk uttaleordbok*, Aschehoug, Oslo.
- Beskow, J., Dahlquist, M., Granström, B., Lundeberg, M., Spens, K.-E. og Öhman, T. 1997, *The Teleface project – disability, feasibility and intelligibility*, in *Proceedings of Fonetik -97, Swedish Phonetics Conference*, Umeå, Sweden, tilgjengelig fra: <<http://www.speech.kth.se/multimodal/papers/fon97teleface.ps.gz>>, [29.04.2010].
- Black, A. W. 2002, *Perfect Synthesis for all of the people all of the time*, [online], IEEE TTS Workshop 2002, Language Technologies Institute Carnegie Mellon University and Cepstral LLC, tilgjengelig fra: <<http://www.cs.cmu.edu/~awb/papers/IEEE2002/allthetime/allthetime.html>>, [20.01.2010].
- Black, A. W. og Lenzo, K. A., *Festvox: The Blizzard Challenge*, [online], Festvox under LTI ved Carnegie Mellon University, tilgjengelig fra: <<http://festvox.org/blizzard/>>, [23.01.2010].
- Boula de Mareüil, P., d'Alessandro, C., Raake, A., Bailly, G., Garcia, M-N. og Morel, M. 2006, *A joint intelligibility evaluation of French text-to-speech synthesis systems: the EvaSy SUS/ACR campaign*, the 5th International Conference on Language Resources and Evaluation, LREC2006, Genoa, Italia, tilgjengelig fra: <<http://hal.inria.fr/docs/00/10/35/71/PDF/LREC06.PDF>>, [22.10.2010].

- Brandt, S. 2005, *Høyere utdanning – tilgjengelig for alle? Studenter med funksjonsnedsettelse og funksjonshemming i høyere utdanning. Kvalitetsreformens betydning og lærestedenes strategier for inkludering*, NIFU STEP, Senter for innovasjonsforskning, tilgjengelig fra: <<http://www.nifustep.no/Norway/Publications/2005/R-2005-04m.pdf>>, [20.10.2010].
- Breivik, T. 2003, *Norsk språkbank – ei ressurs-samling for industri og forskning*, Norsk språkråd, tilgjengelig fra: <<http://www.sprakrad.no/upload/12399/Artikkel-Nordterm2003.doc>>, [16.02.2010].
- Campbell, N. 2007, "Evaluation of Speech Synthesis. From Reading Machines to Talking Machines", s. 29-64, i Dybkjær, L., Hemsén, H. og Minker, W., *Evaluation of Text and Speech Systems*, Text, Speech and Language Technology, Vol. 37, Springer-Verlag, New York.
- Carlson, R., Granström, B., Lennart, N. og Nord, L. 1992, "The 'listening speed' paradigm for synthesis evaluation", s. 63-66, i Huber, D. (red.), *Fonetik '92, the Sixth Swedish Phonetics Conference held in Gothenburg*, Chalmers University of Technology, Göteborg, tilgjengelig fra: <<http://www.speech.kth.se/prod/publications/files/129.pdf>>, [22.10.2010].
- Carlson, R., Granström, B. og Nord, L. 1990, *Evaluation and Development of the KTH Text-to-Speech System on the Segmental Level*, Speech Communication, Vol. 9, Issue 4, s. 271-277, tilgjengelig fra: <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V1C-48V21K0-FM&_user=10&_coverDate=08%2F31%2F1990&_alid=1507997160&_rdoc=1&_fmt=high&_orig=search&_origin=search&_zone=rslt_list_item&_cdi=5671&_sort=r&_st=13&_docanchor=&view=c&_ct=3&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=6bd24e03150bd15f48bc57d1ff6f7c84&searchtype=a>, [29.04.2010].
- Cryer, H. og Home, S. 2010, *Review of methods for evaluating synthetic speech*, Technical report#8, RNIB Centre for Accessible Information, Birmingham, tilgjengelig fra: <http://www.rnib.org.uk/aboutus/Research/reports/accessibility/Documents/2010_02_Evaluating_synthetic_speech_review.doc>, [07.08.2010].
- Cryer, H. og Home, S. 2009, *User attitudes towards synthetic speech for Talking Books*, Technical report#8, RNIB Centre for Accessible Information, Birmingham, tilgjengelig fra: <http://www.rnib.org.uk/aboutus/Research/reports/accessibility/Documents/2009_05_User_attitudes_synthetic_talking_books.doc>, [23.10.2010].
- Cryer, H. og Home, S. 2008, *Exploring the use of synthetic speech by blind and partially sighted people*, Literature review#2, RNIB Centre for Accessible Information, Birmingham, tilgjengelig fra:

- <http://www.rnib.org.uk/aboutus/Research/reports/accessibility/Documents/2008_11_Exploring_use_of_synthetic_speech.doc>, [24.10.2010].
- Dahlberg, A. 2009, *Fagartikkel om dysleksi vår 2009*, Høgskolen i Østfold, Avdeling for lærerutdanning, tilgjengelig fra: <<http://www.stud.hiof.no/~alexsand/fagartikkel.pdf>>, [20.07.2010].
- Dolphin EasyReader – Norwegian Manual for version 3.01*, Dolphin Computer Access AB, tilgjengelig fra: <<http://www.yourdolphin.com/manuals/047EMANT301.pdf>>, [09.10.2010].
- Dybkjær, L., Hemsén, H. og Minker, W. (red.) 2008, *Evaluation of Text and Speech Systems*, Text, Speech and Language Technology, Vol. 37, Springer-Verlag, New York.
- Dynastat, *Speech Intelligibility. Summary of Speech Intelligibility Testing Methods*, [online] Dynastat, Inc., tilgjengelig fra: <<http://www.dynastat.com/Speech%20Intelligibility.htm>>, [29.04.2010].
- Eide, S., Grelland, H., Kristiansen, A., Sævareid, H. I. og Aasland, D. 2003, *Fordi vi er mennesker. En bok om samarbeidets etikk*, 3. utg. 2009, Fagbokforlaget, Bergen.
- Ejerhed, E. og Källgren, G. 2006, *Stockholm Umeå Corpus Version 2.0, SUC 2.0*, [online], Stockholms universitet, Institutionen för lingvistik, tilgjengelig fra: <http://www2.ling.su.se/DaLi/suc/suc2.0_info.html>, [24.04.2010].
- Emerson – a cross-platform Epub and DAISY reader*, [online], distribuert via Google Project Hosting, tilgjengelig fra: <<http://code.google.com/p/emerson-reader/>>, [08.09.2010].
- Endresen, R. T, Simonsen, H. G. og Sveen, A. (red.) 1996, *Innføring i lingvistik*, 3. utg. 2003, Universitetsforlaget, Oslo.
- Engh, Ø. og Ellefsen, L. T. 2010, *Årsrapport Norsk lyd- og blindeskriftbibliotek 2009*, Norsk lyd- og blindeskriftbibliotek, tilgjengelig fra: <<http://www.nlb.no/file.axd?fileDataID=22533cb8-97a0-4589-bdd5-f54a8345f816>>, [25.08.2010].
- Engh, Ø. og Ellefsen, L. T. 2009, *Årsrapport 2008*, Norsk lyd- og blindeskriftbibliotek, tilgjengelig fra: <http://www.nsd.uib.no/polsys/data/filer/aarsmeldinger/AN_2008_55610.pdf>, [07.08.2010].
- Engstrand, O. 1993, “Quality judgements by users of text-to-speech synthesis as a handicap aid”, s. 35-44, i Dufberg, M. og Engstrand, O. 1993, *Perilus: Experiments in speech processes*, Stockholm University, Department of Linguistics, tilgjengelig fra: <<http://su.diva-portal.org/smash/get/diva2:322896/FULLTEXT01>>, [30.09.2010].
- Ericsson, C., Klein, J., Sjölander, K. og Sönnebo, L. 2007, *Filibuster – a new Swedish text-to-speech system*, Talboks- och Punktskriftsbiblioteket, tilgjengelig fra:

- <http://www.speech.kth.se/prod/publications/files/qpsr/2007/2007_50_1_033-036.pdf>, [12.01.2010].
- Faarlund, J. T. 2005, *Revolusjon i lingvistikken. Noam Chomskys språkteori*, Det Norske Samlaget, Gjøvik.
- Faarlund, J. T., Lie, S. og Vannebo, K. I. 1997, *Norsk referansegrammatikk*, 3. utg. 2002, Universitetsforlaget, Oslo.
- Fahlstedt, E. 2004, *SVT Talande textremsa. En användarutvärdering av två talsynteser i en ny tillämpning*, Examensarbete i talteknologi, Institutionen för tal, musikk och hörsel, Kungliga Tekniska Högskolan, tilgjengelig fra: <<http://www.speech.kth.se/prod/publications/files/1665.pdf>>, [23.10.2010].
- FONEMA. *Prosjektbeskrivelse*, [online], prosjektdel av KUNSTI-programmet, opprettet av Norges Forskningsråd for perioden 01.07.02-31.12.06, tilgjengelig fra <<http://www.iet.ntnu.no/projects/fonema/index.php>>, [22.04.2010].
- Forsbom, E. og Wilhelmsson, 2008, *Revision of Part-of-Speech Tagging in Stockholm Umeå Corpus 2.0*, Proceedings of the 3rd Swedish Language Technology Conference, SLTC 2010, s. 39-40, Linköping, tilgjengelig fra: <<http://www.ida.liu.se/conferences/slct2010/slct-proceeding-final.pdf>>, [04.11.2010].
- Francis, A. L., Nusbaum, H. C. og Fenn, K. 2007, *Effects of Training on the Acoustic-Phonetic Representation of Synthetic Speech*, Journal of Speech, Language, and Hearing Research, Vol. 50, Nr. 6, s. 1445-1465, tilgjengelig fra: <<http://jslhr.asha.org/cgi/content/full/50/6/1445>>, [23.10.2010].
- Frawley, W. 1992, *Linguistic Semantics*, Lawrence Erlbaum Associates, New Jersey.
- Fuglerud, K. S. og Solheim, I. 2008, *Synshemmedes IKT-barrierer. Resultater fra undersøkelse om IKT-bruk blant synshemmede*, Norsk Regnesentral, tilgjengelig fra: <<http://www.epractice.eu/files/media/media2141.pdf>>, [23.04.2010].
- Fuglseth, K. og Skogen (red.), K. 2006, *Masteroppgaven i pedagogikk og spesialpedagogikk*, 2. utg. 2007, Cappelen, Oslo.
- Furui, S. 2007, "Speech and Speaker Recognition Evaluation", s. 1-27, i Dybkjær, L., Hensen, H. og Minker, W., *Evaluation of Text and Speech Systems*, Text, Speech and Language Technology, Vol. 37, Springer-Verlag, New York.
- Garcia, M.-N., d'Alessandro, C., Bailly, G., Boula de Mareuil, P. og Morel, M. 2006, *A joint prosody evaluation of French text-to-speech synthesis systems*, LREC 2006 Italy, The International Conference on Language Resources and Evaluation, tilgjengelig fra: <http://hal.inria.fr/docs/00/10/35/57/PDF/LREC06_EVASY_PROSODIE_v4.pdf>, [29.04.2010].

- Gibbon, D., Moore, R. og Winski, R. (red.) 1997, *Handbook of Standards and Resources for Spoken Language Systems. Spoken Language System Assessment*, Mouton de Gruyter, Berlin.
- Goldstein, M. 1995, *Classification of Methods Used for Assessment of Text-to-Speech Systems According to the Demands Placed on the Listener*, Speech Communication, Vol. 16, Issue 3, s. 225-244, tilgjengelig fra: <<http://portal.acm.org/citation.cfm?id=205263>>, [28.04.2010].
- Guevara, E. 2010, *NoWaC: a large web-based corpus for Norwegian*, Proceedings of the NAACL HLT, Sixth Web as Corpus Workshop, s. 1-7, tilgjengelig fra: <<http://www.aclweb.org/anthology/W/W10/W10-1501.pdf>>, [08.11.2010].
- Hals, P. 2008, *Språkteknologisk samarbeid NLB-TPB. Filisynk og Filibuster*, Norsk lyd- og blindeskriftbibliotek, Oslo.
- Hampshire, B., Rüden, J., Carlson, R. og Granström, B. 1982, *Evaluation of centrally produced and distributed synthetic speech*, Quarterly Progress and Status Report, Dept. for Speech, Music and Hearing, s. 18-23, tilgjengelig fra: <http://www.speech.kth.se/prod/publications/files/qpsr/1982/1982_23_2-3_018-023.pdf>, [23.10.2010].
- Hansen, I. L. S. 2008, *IKT og funksjonshemmede. Et potensial for arbeids- og samfunnsliv?*, Fafo, tilgjengelig fra: <<http://www.fafo.no/pub/rapp/10065-10066-10067/10065.pdf>>, [24.04.2010].
- Henton, C. 1992, *Sex and speech synthesis: techniques, successes, and challenges*, Proceedings of the Fourth Australian International Conference on Speech Science and Technology, SST-92, Brisbane, s. 738-743, tilgjengelig fra: <<http://www.assta.org/sst/SST-92/cache/SST-92-SpeakerCharacteristics-p2.pdf>>, [22.10.2010].
- IRTI eCclipseReader-EyeNet DAISY / NISO Talking Book and Text Document Reader*, [online], Innovative Rehabilitation Technology, Inc., tilgjengelig fra: <<http://www.irti.net/eclipse/eCclipseReader/index.html#Spell%20Word>>, [09.10.2010].
- Jekosch, U. 2005, *Voice and Speech Quality Perception. Assessment and Evaluation*, Springer, Heidelberg.
- Jekosch, U. 1993, *Speech Quality Assessment and Evaluation*, EUROSPEECH '93, Third European Conference on Speech Communication and Technology, s. 1387-1394, tilgjengelig fra: <http://www.isca-speech.org/archive/archive_papers/eurospeech_1993/e93_1387.pdf>, [28.04.2010].
- Jekosch, U. 1992, *The Cluster-Identification Test*, ICSLP-1992, Second International Conference on Spoken Language Processing, s. 205-208, tilgjengelig fra:

- <http://www.isca-speech.org/archive/archive_papers/icslp_1992/i92_0205.pdf>, [28.04.2010].
- Johannessen, J. B., Priestley, J., Hagen, K. og Åfarli, T. A. 2009, "The Nordic Dialect Corpus – an advanced research tool", i Jokinen, K. og Bick, E. (red.), *Proceedings of NODALIDA Conference 2009*, s. 73-80, tilgjengelig fra: <<http://dspace.utlib.ee/dspace/bitstream/10062/9730/1/paper26.pdf>>, [08.10.2010].
- Johannessen, J. B. 2003, "Innsamling av språklige data: Informanter, introspeksjon og korpus", i Johannessen, J. B. (red.), Erlenkamp, S., Faarlund, J. T., Lanza, E. og Vonen, A. M., *På språkjakt – problemer og utfordringer i språkvitenskapelig datainnsamling*, Unipub, Oslo.
- Kalvik, M-L., Mihkla, M., Kiissel, I. og Hein, I. 2010, "Estonian: Some Findings for Modelling Speech Rhythmicity and Perception of Speech Rate", i Sojka, P., Horák, A., Kopecek, I. og Pala, K. (red.), *Lecture Notes in Computer Science*, Proceedings of Text, Speech and Dialogue, 13th International Conference, Brno, s. 314-231, tilgjengelig fra: <<http://www.springerlink.com/content/18n1845n172355t6/>>, [24.10.2010].
- Kamstrup, G. W., Mjøvik, E., Rygvold, A-L. og Saltnes, B. G. 2002, *MULTIFUNK: Bringing computer-supported reading one step further*, Norsk Regnesentral, tilgjengelig fra: <<http://publications.nr.no/Multifunk-NR-rapport.pdf>>, [22.10.2010].
- Karlsen, A. V. 2007, *Kommunikasjon via e-post med grafiske symboler 2004-2006. Støttet av IT Funk "IT for funksjonshemmede"*, Høgskolen i Vestfold, tilgjengelig fra: <http://www.tks2.no/ask2/tegnsystemer/grafiske/index_assets/karlsen2007.pdf>, [24.04.2010].
- Kaur, M. og Singh, G. 2010, *Pitch Normalization in Concatenative Speech Synthesis*, ICUTIT2010, tilgjengelig fra: <<http://conf.pcte.edu.in/ICUTIT2010/ResearchPapers/Pitch%20Normalization%20in%20Concatenative%20Speech%20Synthesis.pdf>>, [29.04.2010].
- Kawamura, H. 2006, *DAISY: a better way to read, a better way to publish – a contribution of libraries serving persons with print disabilities*, World Library and Information Congress: 72nd IFLA General Conference and Council, Seoul, Korea, tilgjengelig fra: <<http://archive.ifla.org/IV/ifla72/papers/091-Kawamura-en.pdf>>, [24.04.2010].
- Keller, E. 2000, "Quality improvement of (wireless) phone-based teleservices using advanced speech synthesis techniques", i Pellandini, F., Ansong, M. og Gigueiras-Vidal, A. (red.), *Proceedings of the International Workshop on Intelligent Communication Technologies and Applications, with Emphasis on Mobile Communications*, s. 358-363, tilgjengelig fra: <<http://www.mindsight.cc/cv/pdf.files/Keller-99-COST254-Neuchatel.PDF>>, [23.10.2010].

- Kerscher, G., Kahl, K. og Sutton, J. 2008, *DAISY OK Requirements for Reading Systems and Books*, [online], The DAISY Consortium, tilgjengelig fra:
<<http://www.daisy.org/projects/daisyok/daisy-ok-spec-draft.html?q=project/daisyok>>, [09.10.2010].
- King, M. 2007, ”General Principles of User-Oriented Evaluation”, i Dybkjær, L., Hemsén, H. og Minker, W. (red.), *Evaluation of Text and Speech Systems*, Text, Speech and Language Technology, Vol. 37, 2008, s. 125-161, Springer-Verlag, New York.
- Klarspråk: Språket i lover og andre juridiske tekster*, [online], Norsk språkråd, tilgjengelig fra: <<http://www.sprakrad.no/Klarsprak/Verktoy/Skriverad/Spraket-i-lover-og-andre-juridiske-tekster/>>, [01.09.2010].
- Klatt, D. og Klatt, L. 1989, *Analysis, synthesis, and perception of voice quality variations among female and male talkers*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, s. 820-857, tilgjengelig fra:
<http://www.fon.hum.uva.nl/david/ma_ssp/2010/Klatt-1990-JAS000820.pdf>, [23.10.2010].
- Kristensen, L. K., Gjerde, A. og Knarlag, K. 2008, *Nasjonal pådriver for tilgjengelighet i høyere utdanning. Prosjektrapport. Universell utforming og tilgjengeliggjøring av studielitteratur*, Norsk lyd- og blindeskriftbibliotek og Nasjonal pådriver, tilgjengelig fra: <<http://www.nlb.no/file.axd?fileDataID=a0e18fa3-5412-4f3f-9e74-b760df469fc8>>, [06.11.2010].
- Kominek, J. og Black, A. W. 2003, *CMU ARCTIC databases for speech synthesis*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Technical Report CMU-LTI-03-177, tilgjengelig fra:
<http://festvox.org/cmu_arctic/cmu_arctic_report.pdf>, [03.09.2010].
- Koul, R. 2003, *Synthetic Speech Perception in Individuals With and Without Disabilities*, Augmentative and Alternative Communication, Vol. 19, Nr. 1, s. 49-58, tilgjengelig fra: <<http://informahealthcare.com/doi/abs/10.1080/0743461031000073092>>, [24.10.2010].
- Kristoffersen, K. E., Simonsen, H. G. og Sveen, A. (red.) 2005, *Språk. En grunnbok*, 2. utg. 2008, Universitetsforlaget, Oslo.
- Kvale, K. 2006, *Norsk taleteknologi – statusrapport hausten 2006*, Telenor, tilgjengelig fra: <http://www.telenor.com/rd/pub/rep06/r_12_06.pdf>, [23.04.2010].
- Kulbrandstad, L. A. 1993, *Språkets mønstre. Grammatiske begreper og metoder*, 4. utg., Universitetsforlaget, Oslo.
- Kunnskapsdepartementet 2007, *Rapport fra arbeidsgruppe – tilgjengelig og tilrettelagt faglitteratur for studenter med syns- og lesehemninger*, [online], tilgjengelig fra

- <<http://www.regjeringen.no/nr/dep/kd/dok/hoeringer/hoeringsdok/2007/horing---tilgjengelig-og-tilrettelagt-fa.html?id=465525>>, [22.10.2010].
- Lampert, A. 2004, *Evaluation of the MU-TALK Speech Synthesis System*, ICT Centre, Information & communication technologies, tilgjengelig fra: <<http://www.ict.csiro.au/staff/andrew.lampert/writing/SynthesisEvaluation.pdf>>, [24.10.2010].
- Leith, L. 2006, *Reading the DAISY way*, [online], The DAISY Consortium, tilgjengelig fra: <<http://www.digitaltalkingbook.com/publications/docs/20070315155100/intro-article1.html>>, [25.01.2010].
- Lemmetty, S. 1999, *Review of Speech Synthesis Technology. Master thesis*, Helsinki University of Technology, Department of Electrical and Communications Engineering, tilgjengelig fra: <http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/thesis.pdf>, [27.03.2010].
- Lewin, C. 2000, *Exploring the effects of talking book software in UK primary classrooms*, [online], Journal of Research in Reading, Vol. 23, Nr. 2, s. 149-157, tilgjengelig fra: <<http://onlinelibrary.wiley.com/doi/10.1111/1467-9817.00111/abstract>>, [23.10.2010].
- Léwy, N. og Hornstein, T. 1994, *Text-to-Speech Technology: A Survey of German Speech Synthesis Systems*, UNILAB Technical Report 94.10.2, tilgjengelig fra: <http://www.ubilab.org/publications/print_versions/pdf/germtts.pdf>, [23.10.2010].
- Li, J., Sityaev, D. og Hao, J. 2007, *Sentence Level Intelligibility Evaluation for Mandarin Text-to-Speech Systems Using Semantically Unpredictable Sentences*, INTERSPEECH-2007, s.1350-1353, tilgjengelig fra: <http://www.toshiba-europe.com/research/crl/stg/pdfs/interspeech07_li.pdf>, [04.11.2010].
- Loquendo – Emotional TTS Voices*, [online], Loquendo Vocal Technology and Services, tilgjengelig fra: <http://www.loquendo.com/en/technology/emotional_tts.htm>, [19.05.2010].
- Lydbokavtalen 2008, Avtale om eksemplarfremstilling for funksjonshemmede i medhold av § 17a i lov om opphavsrett til åndsverk m.v. (åndsverkloven)*, tilgjengelig fra: <<http://www.regjeringen.no/upload/KKD/Medier/Lydbokavtalen260608.pdf>>, [04.11.2010].
- Lødrup, P. 1990, *Arverett*, 4. utg. 1999, Peter Lødrup, Institutt for privatrett, Oslo.
- Maegaard, B., Fenstad, J-E., Ahrenberg, L., Kvale, K., Mühlenbock, K. og Heid, B-E. 2007, *KUNSTI – Knowledge Generation for Norwegian Language Technology*, Forskningsrådet, tilgjengelig fra:

- <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.128.5120&rep=rep1&type=pdf>>, [24.04.2010].
- Mariniak, A. 1993, *A Global Framework for the Assessment of Synthetic Speech Without Subjects*, EUROSPEECH '93, Third European Conference on Speech Communication and Technology, s. 1683-1686, tilgjengelig fra: <http://www.isca-speech.org/archive/archive_papers/eurospeech_1993/e93_1683.pdf>, [28.04.2010].
- Meen, D., Svendsen, T. og Natvig, J. E. 2005, *Improving Phone Label Alignment Accuracy by Utilizing Voicing Information*, Proceedings of SPECOM-2005, Patras, Hellas, tilgjengelig fra: <http://www.iet.ntnu.no/~torbjorn/Papers/2005_SPECOM.pdf>, [24.04.2010].
- Mertens, T., Schneider, D., Næss, A. B., Svendsen, T. 2009, *Lexicon Adaptation for Subword Speech Recognition*, Proceedings of ASTRU 2009, Merano, Italia, tilgjengelig fra: <<http://ieeexplore.ieee.org/articleSale/Sarticle.jsp?arnumber=5373296>>, [04.11.2010].
- Morton, K. 1991, *Expectations for Assessment Techniques Applied to Speech Synthesis*, Proceedings of the Institute of Acoustics, Vol. 13, Part 2, tilgjengelig fra: <http://www.morton-tatham.co.uk/publications/to_1994/Morton_1991.pdf>, [24.10.2010].
- Mårtenson, P. og Fjeldstad, A. 1993, *Svenska för norrmän*, 3. utg. 2. opplag 2003, Gyldendal, Oslo.
- Neovius, L. og Raghavendra, P. 1993, *Evaluation of comprehension of KTH text-to-speech with "listening speed" paradigm*, STL-QPSR, Vol. 34, nr. 2-3, 1993, s. 21-30, tilgjengelig fra: <http://www.speech.kth.se/prod/publications/files/qpsr/1993/1993_34_2-3_021-030.pdf>, [30.09.2010].
- Nes, M. 2007, *Appraising and Evaluating the Use of DAISY. For Print Disabled Students in Norwegian Primary- and Secondary Education. Master thesis*, Universitetet i Oslo, Institutt for informatikk, tilgjengelig fra: <http://folk.uio.no/menes/Publications_files/Nes_AppraisingAndEvaluatingDAISY-1.pdf>, [27.03.2010].
- Nordgård, T. 1998, *Språkteknologi i Norge*, [online], Språkrådet, tilgjengelig fra: <http://www.sprakrad.no/Toppmeny/Publikasjoner/Spraaknytt/Arkivet/Spraaknytt_1998/Spraaknytt_1998_4/Spraakteknologi_i_Norge/>, [23.04.2010].
- Norsk lyd- og blindeskriftbibliotek 2010a, *Hvem kan låne*, [online], tilgjengelig fra: <<http://www.nlb.no/sitepageview.aspx?sitePageID=1015>>, [16.04.2010].
- Norsk lyd- og blindeskriftbibliotek 2010b, *Kriterier for rett til produksjon av studielitteratur*, [online], tilgjengelig fra: <<http://www.nlb.no/sitepageview.aspx?articleID=22>>, [20.10.2010].

- NOU 2005:8, *Likeverd og tilgjengelighet. Rettslig vern mot diskriminering på grunnlag av nedsatt funksjonsevne. Bedre tilgjengelighet for alle*, tilgjengelig fra: <<http://www.regjeringen.no/Rpub/NOU/20052005/008/PDFS/NOU200520050008000DDDPDFS.pdf>>, [06.11.2010].
- Næss, T. og Støren, L. A. 2006, *Hvem er de nye studentene? Bakgrunn og studievalg*, NIFU STEP, tilgjengelig fra: <<http://www.nifustep.no/Norway/Publications/2006/NIFU%20STEP%20Arbeidsnotat%203-2006.pdf>>, [23.10.2010].
- Nøklestad, A. og Søfteland, Å. 2007, "Tagging a Norwegian Speech Corpus", i Nivre, J., Kaalep, H-J., Muischnek, K. og Koit, M. (red.), *Proceedings of NODALIDA Conference 2007*, s. 245-248, tilgjengelig fra: <<http://dspace.utlib.ee/dspace/bitstream/10062/2582/1/stud-Nklestad-20.pdf>>, [08.11.2010].
- Olaussen, M. og Haugen, E. 2008, *Begrunnelse for valg av innleser til norsk Filibuster*, Norsk lyd- og blindeskriftbibliotek, Oslo.
- Olofsson, Å. 2002, *Synthetic speech and computer aided reading for reading disabled children*, Reading and Writing: An Interdisciplinary Journal, Volum 4, Nr. 2, s. 165-178, tilgjengelig fra <<http://www.springerlink.com/content/j521536n135x2864/>>, [23.10.2010].
- Pavlovic, C., Rossi, M. og Espesser, R. 1990, *Use of the magnitude estimation technique for assessing the performance of text-to-speech synthesis systems*, Journal of the Acoustical Society of America, Vol. 87, Issue 1, s. 373-382, tilgjengelig fra: <<http://scitation.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=JASMAN0000870000001000373000001&idtype=cvips&gifs=yes&ref=no>>, [29.04.2010].
- Persson, M. 2004, *Utvärdering av några text-till-talomvandlare som läshjälpmiddel*, Examensarbete i datorlingvistik, Institutionen för Lingvistik och Filologi, Uppsala Universitet, tilgjengelig fra: <http://stp.lingfil.uu.se/exarb/arch/2004_persson.pdf>, [23.10.2010].
- Pinker, S. 1995, *The Language Instinct*, Penguin Group, London.
- Pisoni, D. og Hunnicutt, S. 1980, *Perceptual Evaluation of MITALK: The MIT Unrestricted Text-to-Speech System*, Acoustics, Speech and Signal Processing, IEEE International Conference on ICASSP '80, Vol. 5, s. 572-575, tilgjengelig fra: <<http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel6%2F8368%2F26350%2F01170888.pdf%3Farnumber%3D1170888&authDecision=-203>>, [29.04.2010].
- Robson, C. 1993, *Real World Research. A Resource for Social Scientists and Practitioner-Researchers*, 2. utg., 2002, Blackwell, Oxford.

- Sentio Research Norge 2008, *Brukerundersøkelsen 2008*, Norsk lyd- og blindeskriftbibliotek, tilgjengelig fra: <<http://www.nlb.no/sitepageview.aspx?articleID=44>>, [24.10.2010].
- Sikveland, R. O., Öttl, A., Amdal, I., Ernestus, M., Svendsen, T. og Edlund, J. 2010, *Spontal-N: A Corpus of Interactional Spoken Norwegian*, Proceedings of LREC-2010, Veletta, Malta, May 2010, tilgjengelig fra: <http://www.lrec-conf.org/proceedings/lrec2010/pdf/314_Paper.pdf>, [04.11.2010].
- Shiga, Y., Hara, Y. og Nitta, T. 1994, *A Novel Segment-Concatenation Algorithm for a Cepstrum-Based Synthesizer*, ICSLP-1994, Third International Conference on Spoken Language Processing, s. 1783-1786, tilgjengelig fra: <http://www.isca-speech.org/archive/archive_papers/icslp_1994/i94_1783.pdf>, [29.04.2010].
- Simpson, A. 2009, *Breathiness differences in male and female speech. Is H1-H2 an appropriate measure?*, Proceedings of FONETIK 2009, Department of Linguistics, Stockholm Universitet, tilgjengelig fra: <http://www2.ling.su.se/fon/fonetik_2009/172%20simpson_fonetik2009.pdf>, [23.10.2010].
- Siniscalchi, S. M., Svendsen, T. og Lee, C-H. 2008, *A Penalized Logistic Regression Approach to Detection Based Phone Classification*, Proceedings of Interspeech 2008, Brisbane, Australia, Sept. 2008, tilgjengelig fra: <http://www.isca-speech.org/archive/archive_papers/interspeech_2008/i08_2390.pdf>, [24.04.2010].
- Skog, O-J. 1998, *Å forklare sosiale fenomener. En regresjonsbasert tilnærming*, 2. utg., 1. oppl. 2004, Gyldendal Akademisk, Oslo.
- Sonntag, G. og Portele, T. 1998, *Comparative evaluation of synthetic prosody with the PURR method*, 5th International Conference on Spoken Language Processing, tilgjengelig fra: <<http://www.shlrc.mq.edu.au/proceedings/icslp98/PDF/AUTHOR/SL980018.PDF>>, [29.04.2010].
- Sonntag, G. og Portele, T. 1997, *A method for prosody evaluation*, Proceedings of SALT workshop, s. 188-194, Sheffield, tilgjengelig fra: <http://storage.sk.uni-bonn.de/publications/gso_salt97.ps.gz>, [29.04.2010].
- Sperberg-McQueen, C. M. og Burnard, L. 2004, *Language Corpora*, [online], TEI Consortium, tilgjengelig fra: <<http://www.tei-c.org/release/doc/tei-p4-doc/html/CC.html>>, [24.04.2010].
- Språkbanken. *NSTs norske leksikon*, [online], Norsk språkråd (men nå under Nasjonalbiblioteket), tilgjengelig fra: <http://gandalf.aksis.uib.no/sprakbank/ressurser/nokkeltall_norsk.page>, [22.04.2010].
- Ståhl, M. 2009, *Folke vs Henry: En jämförelse av förståelse mellan syntetisk och mänsklig uppläsning av sammanhängande texter*, Stockholms universitet, Institutionen för

- lingvistik och nygrekiska, tilgjengelig fra:
<http://www.uppsatser.se/uppsats/662367c70e/>, [27.03.2010].
- Svendsen, T., Amdal, I., Bjørkan, I., Heggveit, P. O., Meen, D. og Natvig, J. E. 2005, *FONEMA – Tools for realistic speech synthesis in Norwegian*, Proceedings of NORSIG-2005, Stavanger, Sept. 2005, tilgjengelig fra:
http://www.iet.ntnu.no/~torbjorn/Papers/2005_Norsig_Fonema.pdf, [24.04.2010].
- Svennevig, J. 2001, *Språklig samhandling. Innføring I kommunikasjonsteori og diskursanalyse*, 4. utg. 2003, LNU og Cappelen, Oslo.
- Taylor, P. 2009, *Text-to-Speech Synthesis*, Cambridge University Press, Cambridge.
- Tollefsen, M. 2006, *Bruk av syntetisk tale i fulltekst DAISY titler*, [online], MediaLT, tilgjengelig fra: <http://medialt.no/rapport-bruk-av-syntetisk-tale-i-fulltekst-daisy-titler/188.aspx>, [24.04.2010].
- Tomokiyo, L., Peterson, K., Black, A. W. og Lenzo, K. 2006, *Intelligibility of Machine Translation Output in Speech Synthesis*, INTERSPEECH 2006, ICSLP, tilgjengelig fra: <http://www.cs.cmu.edu/~awb/papers/is2006/IS061268.PDF>, [22.08.2010].
- Tännander, C. og Sjölander, K. 2008, *Förstudierapport – Språkteknologiskt samarbete mellan TPB och NLB*, Talboks- och Punktskriftsbiblioteket, Stockholm.
- van Daal, V. og Reitsma, P. 2000, "Computer-assisted learning to read and spell: results from two pilot studies", i *Journal of Research in Reading*, Vol. 23, Nr. 2, s. 181-193, tilgjengelig fra: <http://onlinelibrary.wiley.com/doi/10.1111/1467-9817.00113/abstract>, [23.10.2010].
- van Heuven, V. og Pols, L. 1993, *Analysis and Synthesis of Speech: Strategic Research Towards High-Quality Text-To-Speech Generation*, Mouton De Gruyter, Berlin.
- Venkatagiri, H. 1994, *Effect of sentence length and exposure on the intelligibility of synthesized speech*, *Augmentative and Alternative Communication*, Vol. 10, Nr. 2, s. 96-104, tilgjengelig fra:
<http://informahealthcare.com/doi/abs/10.1080/07434619412331276800?journalCode=aac>, [23.10.2010].
- Wells, J. (25.10.2005), *SAMPA computer readable phonetic alphabet*, [online], Department of Speech, Hearing and Phonetic Sciences, University College London, tilgjengelig fra: <http://www.phon.ucl.ac.uk/home/sampa/>, [06.09.2010].
- "Wikipediaartikkel om Ivar Aasen", (05.08.2010), [online], Wikipedia, tilgjengelig fra:
http://no.wikipedia.org/wiki/Ivar_Aasen, [20.09.2010].
- Wise, B. W., Olson, R. K. og Ring, J. 1997, "Teaching Phonological Awareness With and Without the Computer", s. 254-274, i Hulme, C. og Snowling, M. (red.), *Dyslexia: Biology, Cognition and Intervention*, Whurr, London.

Xu, J., Matusov, E., Zens, R. og Ney, H. 2005, *Integrated Chinese Word Segmentation in Statistical Machine Translation*, Proceedings of IWSLT, s. 141-147, Pittsburgh, tilgjengelig fra: <<http://www.mt-archive.info/IWSLT-2005-Xu.pdf>>, [24.10.2010].

Åfarli, T. A. og Eide, K. M. 2003, *Norsk generativ syntaks*, Novus, Oslo.

Årsand, E., Walseth, O. A., Andersson, N., Finnset, W., Grav, J., Rinde, E., Bergvik, S. og Henriksen, A. 2004, *Alternative brukergrensesnitt for helserelaterte tjenester på mobile terminaler*, Nasjonalt senter for telemedisin, Universitetssykehuset Nord-Norge, tilgjengelig fra: <http://img7.custompublish.com/getfile.php/122511.357/Rapport_Brukergrensesnitt.pdf?return=www.telemed.no>, [23.04.2010].

Vedlegg

1 Vedlegg til brukerundersøkelsen

1.1 Spørreskjema

Spørreskjemaet til brukerundersøkelsen er gjengitt nedenfor. Det elektroniske spørreskjemaet som informantene benyttet seg av, er tilgjengelig via <http://www.nlbphp.net/brageskjema/> ut 2010. Det elektroniske skjemaet er ikke lenger i bruk.

DEL A – Bakgrunnsspørsmål

Først ønsker vi å vite mer om din bakgrunn.

1. Lånernummer

2. Kjønn

- Mann
- Kvinne

3. Alder

4. Morsmål (og hovedmål)

5. Bakgrunn for lånerett

- Blindhet
- Nedsatt synsevne
- Dysleksi
- Annet

6. Har du produksjonsrett?

- Ja
- Nei, jeg har kun lånerett

7. Bosted (fylke)

- Østfold
- Akershus
- Oslo
- Hedmark
- Oppland
- Buskerud
- Vestfold
- Telemark
- Aust-Agder
- Vest-Agder
- Rogaland
- Hordaland
- Sogn og Fjordane
- Møre og Romsdal
- Sør-Trøndelag
- Nord-Trøndelag
- Nordland
- Troms
- Finnmark
- Svalbard

8. Hvor lenge har du vært låner hos NLB?

- Mindre enn 1 år
- Fra 1 til 3 år
- Fra 4 til 6 år
- Fra 7 til 9 år
- 10 år eller lenger

9. Hvilket lærested studerte du ved vårsemesteret 2010?

10. Hvilket studieprogram deltok du i vårsemesteret 2010?

11. Hvilke av disse norske talesyntesene har du kjennskap til?

- Ingen kjennskap til talesyntese
- Acapelas Kari
- Acapelas Olav

- Acapelas Trygve
- Acapelas Vegard
- NLBs Brage
- Nuances Nora
- Nuances Stine
- Lingspeak Arne
- Lingit Isak HTS
- Telenor Talsmann
- eSpeak norsk
- Andre

12. Hvilke av disse norske talesyntesene har du benyttet deg av?

- Ingen kjennskap til talesyntese
- Acapelas Kari
- Acapelas Olav
- Acapelas Trygve
- Acapelas Vegard
- NLBs Brage
- Nuances Nora
- Nuances Stine
- Lingspeak Arne
- Lingit Isak HTS
- Telenor Talsmann
- eSpeak norsk
- Andre

13. Hvorfor har du benyttet deg av nettopp disse talesyntesene?

14. Hvor lenge har du benyttet deg av en norsk talesyntese?

- Har ikke benyttet meg av talesyntese
- Mindre enn 1 år
- Fra 1 til 3 år
- Fra 4 til 6 år
- Fra 7 til 9 år
- 10 år eller lenger

15. Dersom du kunne velge en av disse norske talesyntesene som oppleser for din studielitteratur, hvilken ville du ha foretrukket?

- Foretrekker ikke talesyntese
- Acapelas Kari
- Acapelas Olav
- Acapelas Trygve

- Acapelas Vegard
- NLBs Brage
- Nuances Nora
- Nuances Stine
- Lingspeak Arne
- Lingit Isak HTS
- Telenor Talsmann
- eSpeak norsk
- andre

16. Grunngi svaret i spørsmål 15.

17. Har du i dette eller tidligere semestre benyttet deg av studiemateriell fra NLB, som er produsert med andre talesynteser enn Brage?

- Ja
- Nei

DEL B - Stemmekvalitet og Brage

I denne delen vil vi vite hva du synes om selve stemmen til Brage.

18. Hvor godt liker du selve stemmen til Brage?

- Liker stemmen godt
- Liker stemmen nokså godt
- Verken liker eller misliker stemmen
- Liker stemmen mindre godt
- Liker ikke stemmen i det hele tatt

19. Med begrepet toneleie menes frekvensnivået som definerer hvor mørk eller hvor lys en stemme kan være. Dersom du kunne velge toneleie til en syntetisk stemme, hvilket ville du ha foretrukket?

- Mannsstemme bass (mørk)
- Mannsstemme tenor (lys)
- Kvinnestemme alt (mørk)
- Kvinnestemme sopran (lys)
- Spiller ingen rolle

20. Hva synes du om toneleiet til Brage?

21. Prosodi omfatter ”melodiske” sider ved talespråket, som setningsmelodi (variasjon innen intonasjon), tonelag (forskjellen på ”bønder” og ”bønner”), trykk og rytme. Hvilke oppfatninger har du gjort deg om prosodien til Brage?

22. Sammenliknet med menneskelig opplesing, hvor naturlig synes du Brage høres ut?

- Naturlig
- Nokså naturlig
- Verken naturlig eller unaturlig
- Nokså unaturlig
- Unaturlig

23. Hvor viktig er det for deg at en talesyntese, som skal produsere studielitteratur, likner menneskelig opplesing?

- Viktig
- Nokså viktig
- Verken viktig eller uviktig
- Nokså uviktig
- Uviktig

24. Dersom du kunne velge mellom å få studielitteratur med syntetisk tale eller menneskelig tale, hva ville du ha foretrukket?

- Menneskelig tale
- Syntetisk tale
- Begge deler
- Spiller ingen rolle

25. Hva synes du om å få produsert studielitteratur med syntetisk tale?

DEL C - Uttale og Brage

Spørsmålene i denne delen omfatter dine erfaringer med Brages uttale knyttet til studielitteratur.

26. Hvordan opplever du den generelle artikuleringa til Brage?

- Tydelig
- Nokså tydelig
- Nokså utydelig
- Utydelig

27. Hvor ofte erfarer du at Brage uttaler et ord feil, slik at du ikke forstår ordet?

- Ofte
- Av og til
- Sjelden
- Aldri

28. Hvor ofte erfarer du at Brage uttaler et ord feil, slik at du ikke forstår setninga?

- Ofte
- Av og til
- Sjelden
- Aldri

29. Hvor ofte opplever du at deler av et ord kuttet eller slukes?

- Ofte
- Av og til
- Sjelden
- Aldri

**30. Eksempler på et sammensatt ord kan være "terminologiliste" og "eksamensvakt".
Hvor ofte opplever du at sammensatte ord uttales feil?**

- Ofte
- Av og til
- Sjelden
- Aldri

31. Homografer er ord som skrives likt, men som har ulik betydning. Eksempler på homografpar kan være substantivet "landet" og verbet "landet", eller substantivet "budsjetter" og verbet "budsjetter". Hvor ofte opplever du at en homograf betones feil?

- Ofte
- Av og til
- Sjelden
- Aldri

32. Dato, valuta, matematiske og kjemiske formler, telefonnumre, målenheter m.m., er eksempler på tall og talluttrykk. Hvor ofte erfarer du at tall eller uttrykk med tall uttales uforståelig eller feil?

- Ofte
- Av og til
- Sjelden
- Aldri

33. Eksempler på norske egennavn kan være "Petter Northug" og "Svolvær". Hvor ofte opplever du at uttalen av norske egennavn er uforståelig eller feil?

- Ofte
- Av og til
- Sjelden
- Aldri

34. Eksempler på utenlandske egennavn kan være "Barack Obama" og "Düsseldorf". Hvor ofte opplever du at uttalen av utenlandske egennavn er uforståelig eller feil?

- Ofte
- Av og til
- Sjelden
- Aldri

35. Hvordan synes du Brage mestrer innslag av nynorsk?

- Godt
- Nokså godt
- Ok
- Nokså dårlig
- Dårlig
- Ingen innslag av nynorsk

36. Hvordan synes du Brage mestrer innslag av engelsk?

- Godt
- Nokså godt
- Ok
- Nokså dårlig
- Dårlig
- Ingen innslag av engelsk

37. Hvordan synes du Brage mestrer faguttrykk innenfor studieretninga di?

- Godt

- Nokså godt
- Ok
- Nokså dårlig
- Dårlig

DEL D - Brage som formidler

Spørsmålene i denne delen er avgrenset til dine erfaringer med hvordan Brage formidler studielitteratur.

38. Hvor velegnet mener du selve stemmen til Brage er til å formidle sjangeren studielitteratur?

- Velegnet
- Noe velegnet
- Verken velegnet eller uegnet
- Noe uegnet
- Uegnet

39. Hvordan opplever du å lytte til studielitteratur, produsert med Brage, over tid?

- Behagelig
- Nokså behagelig
- Ok
- Nokså ubehagelig
- Ubehagelig

40. Opplever du at du må anstrenge deg mer for å holde konsentrasjonen oppe når du lytter til lydbøker, som er produsert med Brage, sammenliknet med lydbøker, som er produsert med menneskelig innlesing?

- Ja
- Nei
- Opplever ingen forskjell

41. Hvordan synes du det er å få med seg meningsinnholdet i en tekst, som er produsert med Brage?

- Lett
- Nokså lett
- Verken lett eller vanskelig
- Nokså vanskelig

- Vanskelig

42. Hvordan opplever du det er å få med deg meningsinnholdet i en tekst, som er produsert med Brage, sammenliknet med en tekst, som er produsert med innlest tale?

- Lettere
- Noe lettere
- Noe vanskeligere
- Vanskeligere
- Opplever ingen forskjell

43. Hva synes du om lesetempoet til Brage i normalhastighet?

- Langsomt
- Nokså langsomt
- Passelig
- Nokså raskt
- Raskt

44. Hvilken oppfatning har du gjort deg om flyten i lesinga til Brage?

45. Pauser fungerer som overgangsmarkører i tekster, produsert med talesyntese. En pause kan for eksempel markere overgang fra ei overskrift, ei setning, et avsnitt eller et sideskift. Hvordan opplever du mengden av pauser i studielitteratur, som er produsert med Brage?

- For mange
- Nokså mange
- Passelig
- Nokså få
- For få

46. Hvordan opplever du lengden på pausene i studielitteratur, som er produsert med Brage?

- For lange
- Nokså lange
- Passelig
- Nokså korte
- For korte

47. Hvilken grad av tillit har du til Brage som formidler av pensumbøkene dine?

- Stor tillit

- Delvis tillit
- Liten tillit
- Ingen tillit

DEL E - Dine lesevaner

Spørsmålene i denne delen er knyttet til dine generelle lesevaner og erfaringer med å lese studielitteratur med Brage.

48. Hva slags avspillingsutstyr benytter du deg av når du lytter til studielitteratur?

49. Om lag hvor mange lydbøker leser du i måneden, uten å inkludere studielitteratur?

- Ingen
- 1-2
- 3-4
- 5-6
- Mer enn 6

50. Hva slags type litteratur låner du på lydbok ved sida av studielitteratur?

- Låner kun studielitteratur
- Skjønnlitteratur (noveller, romaner, lyrikk m.m.)
- Tidsskrifter og magasiner
- Aftenposten
- TV- og Radioguiden
- Sakprosa (biografier, faktabøker m.m.)
- Faglitteratur som ikke er relatert til dine studier

51. Om lag hvor mange innleste studie bøker har du lest i løpet av din studieperiode?

- Ingen
- 1-4
- 5-9
- 10-14
- 15 eller flere

52. Om lag hvor mange studie bøker med Brage har du lest?

- Ingen

- 1-2
- 3-4
- 5-6
- Mer enn 6

53. Hvor lenge hører du vanligvis sammenhengende på studielitteratur, produsert med Brage?

- 1 til 14 minutter
- 15 til 29 minutter
- 30 til 44 minutter
- 45 til 59 minutter
- 60 minutter eller lenger

54. Om lag hvor mye tid per uke bruker du gjennomsnittlig til å lese studielitteratur? Ta utgangspunkt i vårsemesteret 2010.

- 1-9 timer
- 10-19 timer
- 20-29 timer
- 30 timer eller mer

55. Hvor viktig er det for deg å kunne lese raskt og ha muligheten til å skimlese?

- Viktig
- Nokså viktig
- Verken viktig eller uviktig
- Nokså uviktig
- Uviktig

56. Tilpasser du avspillingshastigheten på din DAISY-spiller? Hvis nei, gå direkte til spørsmål 60.

- Ja
- Nei

57. Hvorfor tilpasser du avspillingshastigheten?

58. Hvor velegnet synes du Brage er til å lese i en høyere avspillingshastighet?

- Velegnet
- Nokså velegnet
- Nokså uegnet
- Uegnet

Avslutning

Helt til slutt ønsker vi å stille noen generelle spørsmål om Brage knyttet til NLBs produksjon.

59. Synes du at du mottar studiebøker, som er produsert med Brage, innen rimelig tid?

- Ja
- Nei

60. Uavhengig av om du har eller skulle få produksjonsrett, hvor lang tid forventer du at det skal ta å produsere ei gjennomsnittlig stor pensumbok (ca. 400 sider), regnet fra datoen du bestiller boka hos NLB og fram til dagen du mottar den?

- 1 til 2 uker
- 3 til 4 uker
- 5 til 6 uker
- Mer enn 6 uker

61. En skjermleser er et program som leser opp tekst på en dataskjerm. Bruker du skjermleser?

- Ja
- Nei

62. Kunne du tenke deg å bruke Brage som skjermleser?

- Ja
- Nei
- Bruker ikke skjermleser

63. Talesyntese kan gi mulighet til å utvide litteraturtilbudet. Hvordan stiller du deg til at litteratur, som NLB ellers ikke ville ha prioritert å produsere med menneskelig innlesing, i stedet produseres med Brage?

- Positiv
- Nokså positiv
- Verken positiv eller negativ
- Nokså negativ
- Negativ

64. Hvordan opplever du tilgangen til relevant faglitteratur?

65. Har du andre kommentarer vedrørende produksjoner med Brage, som du ikke har fått meddelt gjennom spørreskjemaet?

66. Kan du tenke deg å delta i intervju?

Sett kryss dersom du ønsker å delta i intervju.

2 Vedlegg til SUS-test

2.1 Intransitive setninger

1. En tråd levde over en sint hjelp.
2. Et rom fantes mot et stort brev.
3. Et lån reiste fra et fritt skilt.
4. En dom lå etter et langt bord.
5. En kreft reiste hos en høy sjel.
6. En kveld bidro uten en hel død.
7. En frykt fantes gjennom en bred sønn.
8. Et barn ba mot et trygt gram.
9. Et valg bodde bak et trygt lys.
10. En natt spilte på en full grad.
11. En ferd tenkte uten en from kreft.
12. Et rom vokste med en flott grad.

2.2 Transitive setninger

1. En fast sak svarte et hjem.
2. En glad kveld ga en start.
3. En lang jobb gikk et år.
4. Et hardt plan fikk en plass.
5. En frisk sikt trengte en bønn.
6. En syk kunst følte et poeng.
7. En fast ånd prøvde et vern.
8. En død stein lærte en part.
9. En sint jobb skapte et liv.
10. Et løst blod grep en stat.
11. En trygg tekst endte et fag.
12. En redd mat ansatte en sum.

2.3 Imperative setninger

1. Drift en lov og et grep!
2. Ens et krav og en post!
3. Tvil en stein og en makt!
4. Lik et tre og et kjøp!
5. Tegn en kveld og en post!
6. Spør et vann og en svikt!
7. Vis en rett og et håp!
8. Del en stat og en tråd!
9. Finn et plan og et bånd!
10. Press et svar og en ferd!
11. Tap en gud og en lek!

12. Varm et rom og en plan!

2.4 Spørresetninger

1. Når kunne en lov omfatte en rett slutt?
2. Når måtte et vær plassere en fast stil?
3. Når kunne en lek takle et stilt vern?
4. Når ville et ben leie et greit punkt?
5. Når ville et skilt hente en redd arv?
6. Når ville et smil kjenne et rent hus?
7. Når ville en lek glede en trygg feil?
8. Når skulle en stat begrense et nytt press?
9. Når kunne en tid bryte et tungt fag?
10. Når kunne et kurs selge en rød stein?
11. Når ville en ånd leie en lik bønn?
12. Når kunne et kjøtt hente et fullt berg?

2.5 Relative setninger

1. Et bord fulgte et slag som kjente.
2. En stund brukte en frukt som økte.
3. En rett ansatte en drift som tok.
4. En sang viste et salg som fikk.
5. En hud nevnte en slutt som slo.
6. En gift viste en lek som begrep.
7. Et gull eide en hud som etablerte.
8. En hånd spurte en mann som solgte.
9. En stein svarte en kveld som trengte.
10. Et drap spurte et salt som hørte.
11. Et brev brukte en ro som økte.
12. Et grep trengte et smil som viste.

2.6 Setningsrekkefølgen testen ble gjennomført med

1. Når ville en lek glede en trygg feil?
2. Del en stat og en tråd!
3. Vis en rett og et håp!
4. En rett ansatte en drift som tok.
5. Et bord fulgte et slag som kjente.
6. Et barn ba mot et trygt gram.
7. Spør et vann og en svikt!
8. Når kunne en lov omfatte en rett slutt?
9. Når ville et ben leie et greit punkt?
10. En kveld bidro uten en hel død.
11. Når skulle en stat begrense et nytt press?

12. En trygg tekst endte et fag.
13. Varm et rom og en plan!
14. En død stein lærte en part.
15. En stein svarte en kveld som trengte.
16. Lik et tre og et kjøp!
17. Press et svar og en ferd!
18. Når ville et smil kjenne et rent hus?
19. Når ville et skilt hente en redd arv?
20. Et valg bodde bak et trygt lys.
21. En frisk sikt trengte en bønn.
22. En stund brukte en frukt som økte.
23. Et lån reiste fra et fritt skilt.
24. En fast ånd prøvde et vern.
25. Når kunne en tid bryte et tungt fag?
26. En natt spilte på en full grad.
27. Et rom vokste med en flott grad.
28. En gift viste en lek som begrep.
29. En syk kunst følte et poeng.
30. En glad kveld ga en start.
31. Når ville en ånd leie en lik bønn?
32. Et drap spurte et salt som hørte.
33. Når kunne et kjøtt hente et fullt berg?
34. Tegn en kveld og en post!
35. Et løst blod grep en stat.
36. Et rom fantes mot et stort brev.
37. Finn et plan og et bånd!
38. Tvil en stein og en makt!
39. En fast sak svarte et hjem.
40. Ens et krav og en post!
41. Et gull eide en hud som etablerte.
42. Når måtte et vær plassere en fast stil?
43. Et grep trengte et smil som viste.
44. En dom lå etter et langt bord.
45. En frykt fantes gjennom en bred sønn.
46. En sang viste et salg som fikk.
47. En ferd tenkte uten en from kreft.
48. Når kunne et kurs selge en rød stein?
49. En lang jobb gikk et år.
50. Tap en gud og en lek!
51. En sint jobb skapte et liv.
52. En hud nevnte en slutt som slo.
53. En tråd levde over en sint hjelp.
54. Når kunne en lek takle et stilt vern?
55. En kreft reiste hos en høy sjel.
56. Et hardt plan fikk en plass.
57. Drift en lov og et grep!
58. En hånd spurte en mann som solgte.
59. Et brev brukte en ro som økte.
60. En redd mat ansatte en sum.